

PCT

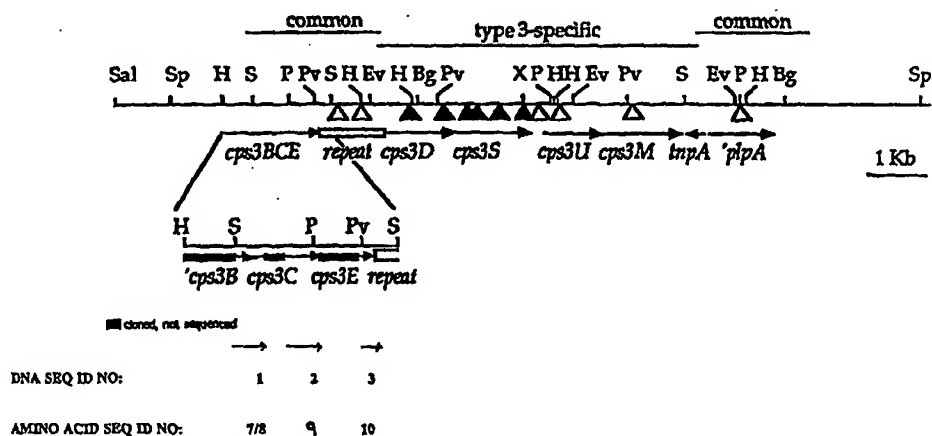
WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : C12N 15/31, C07K 14/31, C12N 15/52, 15/53, 15/60, 15/54, 15/61, C12Q 1/68, C12P 19/04, A61K 39/09, C12N 9/04, 9/88, 9/12, 9/90, 1/21</p>	<p>A1</p>	<p>(11) International Publication Number: <b>WO 95/31548</b> (43) International Publication Date: 23 November 1995 (23.11.95)</p>
<p>(21) International Application Number: PCT/US95/06119 (22) International Filing Date: 16 May 1995 (16.05.95) (30) Priority Data: 08/243,546 16 May 1994 (16.05.94) US (60) Parent Application or Grant (63) Related by Continuation US 08/243,546 (CIP) Filed on 16 May 1994 (16.05.94) (71) Applicant (for all designated States except US): THE UAB RESEARCH FOUNDATION [US/US]; Suite 1120G, 701 South 20th Street, Birmingham, AL 35294-0111 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): YOTHER, Janet [US/US]; 2208 Heatherbrooke Road, Birmingham, AL 35242 (US). DILLARD, Joseph, P. [US/US]; 603 The Lane, Hinsdale, IL 60521 (US).</p>	<p>(74) Agent: PARKER, David, L.; Arnold, White &amp; Durkee, P.O. Box 4433, Houston, TX 77210 (US). (81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p>	

(54) Title: *STREPTOCOCCUS PNEUMONIAE* CAPSULAR POLYSACCHARIDE GENES AND FLANKING REGIONS



(57) Abstract

Disclosed is the identification, cloning and sequencing of flanking DNA regions common to all polysaccharide capsule types in *Streptococcus pneumoniae*. Also disclosed are particular type-specific genes and gene products that direct the formation of the *S. pneumoniae* serotype-specific polysaccharide capsule. Methods are provided for detecting *S. pneumoniae* and for constructing gene cassettes that may be transferred as a unit during transformation and used to direct the expression of specific serotypes of *S. pneumoniae* capsules.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

- 1 -

DESCRIPTIONSTREPTOCOCCUS PNEUMONIAE CAPSULAR  
POLYSACCHARIDE GENES AND FLANKING REGIONS

5

BACKGROUND OF THE INVENTION

10 The present application is a continuation-in-part of  
co-pending U.S. Patent Application Serial No. 08/243,546,  
filed May 16, 1994. The entire text and figures of which  
disclosure is specifically incorporated herein by  
reference without disclaimer. The government owns  
certain rights in the present invention pursuant to grant  
15 number AI28457 from the Public Health Service and  
T32 AI07041-13 from the National Institutes of Health.

1. Field of the Invention

20 The present invention relates generally to the  
fields of bacterial capsule formation and the genes  
responsible for polysaccharide synthesis. More  
particularly, it concerns the genes and gene products  
that direct the formation of the *Streptococcus pneumoniae*  
25 serotype-specific polysaccharide capsule. The present  
invention also includes the identification of non-type  
specific gene sequences, flanking the capsule genes, and  
their use for the directed expression of specific  
serotypes of *S. pneumoniae* capsules.

30

2. Description of the Related Art

35 Infections due to *S. pneumoniae* are among the top  
ten causes of death in the United States. The normal  
populations most affected are young children and the  
elderly: pneumococcal pneumoniae, mainly affecting the  
elderly, causes >40,000 deaths per year among ~500,000

- 2 -

cases and represents 60 to 80% of all bacterial pneumoniae; pneumococcal meningitis, with ~4000 cases/year, represents 11% of the total meningitis cases and has a fatality rate of >30% - greater than twice that of the two other leading causes, *N. meningitidis* and *H. influenzae*; bacteremia, usually following pneumoniae or meningitis, accounts for >35,000 cases per year (>30% fatal); and otitis media, the most frequent reason for pediatric office visits after well-child care, is caused by *S. pneumoniae* in ~50% of cases (ACIP, 1981; ACIP, 1989; Austrian, 1984; Burke et al., 1971; Center for Disease Control, 1978; Johnston and Sell, 1964; Koch and Dennison, 1974).

Other populations have an even higher incidence of pneumococcal infections: approximately 30% of sickle cell children will have severe pneumococcal infections in the first three years of life and ~35% of those will die (Overturf, et al., 1977; Powars, et al., 1981; Powars, 1975); in both adults and children with HIV infections, *S. pneumoniae* is the major cause of invasive bacterial respiratory disease (Janoff et al., 1992). Patients with lymphomas, Hodgkins disease, multiple myeloma, splenectomy, and other debilitating diseases or immunologic deficiencies, are particularly susceptible to serious pneumococcal disease, as are those with chronic illnesses such as diabetes mellitus and heart disease. Furthermore, strains of *S. pneumoniae* are emerging that harbor resistances to multiple antibiotics, including penicillin (Appelbaum, 1992; Jacobs et al., 1978; Landesman et al., 1982).

The polysaccharide capsule of *S. pneumoniae* is the major virulence determinant of this organism. Despite early studies of the genetics, pathogenesis, and immunology of capsular polysaccharides, it remains unclear why certain capsular types appear to have a



- 3 -

greater capacity to cause disease. Of the more than 80 known capsular serotypes, 23 account for more than 90% of all pneumococcal infections.

5           In children, the most prevalent types are 3, 6, 14, 19, and 23, (Gray and Dillon, 1986), whereas in adults types 1, 3, 4, 6, 7, 8, 9, 12, 14, 18, 19, 23 prevail (Finland and Barnes, 1977). In assays of opsono-  
10           phagocytosis (Branconier and Odeberg, 1982; Giebink et al., 1977; Knecht et al., 1970), complement activation and deposition (Fine, 1975; Gordon et al., 1986; Hostetter, 1986; Stephens et al., 1977; Winkelstein et al., 1980; Winkelstein et al., 1976), and mouse virulence (Briles et al., 1992; Briles et al., 1986; Knecht et al.,  
15           1970; MacLeod, 1965; Walter et al., 1941; Yother et al., 1982), levels of virulence have frequently been found to vary with the type of capsule expressed. For example, isolates expressing type 3, 4, and 19 capsules are highly resistant to phagocytosis, whereas those expressing types  
20           6A, 14, 23 and 37 are significantly less resistant (Branconier and Odeberg, 1982; Hostetter, 1986; Knecht et al., 1970; Wood and Smith, 1949).

          The importance of the capsule also results from the  
25           fact that anti-capsular antibodies are highly protective against infection. Nonetheless, the current polysaccharide-based vaccine is not particularly useful in some of the populations most affected by pneumococcal disease, e.g., the very young and the elderly, because of  
30           poor or absent immune response to polysaccharide antigens.

          The ability to produce improved vaccines and  
therapies for pneumococcal infections will most likely be  
35           the result of a better understanding of the basic pathogenic mechanisms of the organism. This understanding necessarily includes the genetic basis for the expression

- 4 -

of serotype-specific polysaccharides and the role of capsular type *per se* in pathogenesis.

Some 85 different serotypes of *Streptococcus pneumoniae*, differing in the structure of the polysaccharide produced, have been identified (van Dam et al., 1990). The basis for the emergence of new capsule types remains obscure. Whether influenced by mutation, recombination, or immune selection, genetic exchange of DNA is likely to have played a major role in the evolution of capsule types. It is known that pneumococcal capsule types can be changed through genetic transformation *in vitro* (Dawson, 1930; Dawson and Sia, 1931; Langvad-Nielson, 1944; Avery et al., 1944). Epidemiological studies suggest that a significant degree of genetic exchange occurs *in vivo* (Crain et al., 1990; Coffey et al., 1991; Versalovic et al., 1993). However, the mechanism by which capsule types are exchanged is not fully understood.

Extensive study was made of the genetics of capsular polysaccharide synthesis in *S. pneumoniae* using spontaneous mutants with defects in biosynthetic functions (Effrussi-Taylor, 1951; Ravin, 1960; Bernheimer and Wermundsen, 1972). The results of these studies indicated that the genes for polysaccharide synthesis were closely linked and could be transferred as a unit during genetic transformation. A cassette-type model of capsule type change based on this data has been proposed (Taylor, 1949; Austrian et al., 1959; Bernheimer and Wermundsen, 1972). According to the model, the type-specific genes for each capsule type would be present only in the genome of a strain of that capsule type and would show little homology to the type-specific genes of other capsule types. The type-specific genes would be located in homologous sites in the different chromosomes, clustered together between regions of highly

- 5 -

homologous flanking DNA. During transformation, recombination would occur in the flanking regions, resulting in the replacement of the recipient's type-specific region by that of the donor.

5

The clustering of capsule biosynthetic genes proposed by the model is analogous to the organization that has been observed in the gram negative bacteria *Escherichia coli* (K antigens) (Roberts et al., 1988),  
10 *Neisseria meningitidis* (Frosch et al., 1989), and *Haemophilus influenzae* (Kroll et al., 1989). For each of these organisms, the type-specific region encoding biosynthetic functions (region 2) is flanked by highly homologous regions necessary for polysaccharide  
15 translocation (region 1) and modification (region 3). Since *H. influenzae*, like *S. pneumoniae*, is naturally transformable, it has been proposed that capsule type change in this pathogen may occur by transformation with the type-specific gene cluster from a different serotype  
20 (Zwahlen et al., 1989).

The one exception to the cassette model of capsule type change in *S. pneumoniae* is binary capsule formation. When non-encapsulated mutants have been transformed with  
25 chromosomal DNA from a strain of a different capsule type, most of the encapsulated transformants express the capsule type of the donor. However, at a frequency 10 to 100 times lower, encapsulated transformants are obtained which express both capsules (Bernheimer and Wermundsen,  
30 1972). In some of these transformants, the second set of capsule genes is closely linked to the original set. However, these strains are unstable, and, at high frequency, lose the ability to produce the original capsule type. In binary strains in which the acquired  
35 capsule genes are unlinked to the original genes, binary capsule production is stable (Bernheimer and Wermundsen, 1969). Elucidation of the mechanism of binary capsule

- 6 -

type formation may be the key to understanding novel capsule type creation in *S. pneumoniae*.

It is clear that a better understanding of the genetics of capsular polysaccharide synthesis in *Streptococcus pneumoniae* is needed. The identification of type-specific capsular genes and the ability to transfer them, singly or as a gene cassette, to desired recipients, will elucidate the role of capsular types in virulence and allow easy identification of *S. pneumoniae* serotype. This ability will improve existing methods of diagnosis, identifying not only the presence of *S. pneumoniae* but also the capsular type of the invading strain. Furthermore, it will allow construction of strains producing elevated levels of capsular polysaccharides for improved vaccines.

#### SUMMARY OF THE INVENTION

##### 20 Capsular Polysaccharide Genes and Flanking Regions

The present invention arises out of the discovery and sequence characterization of a gene family that confers on *S. pneumoniae* the ability to produce type-specific capsules that define the serotype of the organism. The inventors refer to this gene family as the capsule synthesis or *cps* genes. These genes encode the various enzymatic functions of capsule synthesis and determine the particular structure of the capsule polysaccharide that is produced, and thereby define serotype. These genes, designated *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *tnpA*, and *plpA*, map to specific DNA segments of sizes believed to range from about 0.5 kb to greater than 10 kb that appear to be type-specific for *S. pneumoniae*. Based upon the findings of the inventors, many type-specific genes may be distinguished on the

- 7 -

basis of restriction fragment length polymorphism (RFLP) analysis.

The present invention also includes the discovery and sequence characterization of non-type specific DNA regions that flank both sides of the *cps* locus. These flanking DNA segments can be used to identify the location of *cps* flanking DNA from any strain of *S. pneumoniae*. This invention thus provides the ability to identify the *cps* locus within all strains and allows for the subsequent isolation and characterization of all genetic elements involved in determining *S. pneumoniae* serotype.

The classification of *S. pneumoniae* strains is based on serological analysis of cell surface structures. 85 distinct serotypes have been identified to date based on the formation of surface molecules. The formation of the cell surface of *S. pneumoniae*, and in particular its polysaccharide capsule, has, until now, eluded characterization at the molecular genetic level. However, studies of the biosynthesis of the polysaccharide capsule have revealed that at least some of the genes are likely to include enzymes involved in the preparation of the sugar backbones for incorporation into the saccharide backbone, such as UDP-glucose dehydrogenase.

As mentioned above, these polymorphic "type-specific" sequence regions were found to be bounded or flanked by "non-type specific" regions having sequence elements that are apparently shared among the various subtypes. These regions, referred to as the left and right flanking regions, extend for, at least, 1 to 3 kb on either side of the *cps* genes. Thus, in type 3, the entire length of the capsule synthesis genes, including the non-type specific flanking regions, and any DNA

- 8 -

sequences in between, is greater than 9 kb. In other capsule types the length is on the order of about 5 to 20 kb, with the maximum length being related to the complexity of the polysaccharide encoded. Importantly, it is these flanking regions that allow recombination and integration of the type specific capsule genes to occur. Thus, when a selected *cps* gene or genes is positioned between the flanking regions, the resultant construct can be stably integrated into a *S. pneumoniae* host.

10

The present discoveries concerning the *cps* gene regions, the identification of conserved flanking regions, and the construction of erythromycin resistant insertions in adjacent, non-type specific DNA elements, allows for the changing of capsular serotypes by "cassetting-in" the biosynthetic genes for different serotypes. This methodology may be employed for generating high yield capsular polysaccharide producing strains of different (heterologous) serotypes. For a high yielding strain, the existing serotype biosynthetic genes may be deleted and a different serotype's genes inserted ("cassetted in"). These other serotypes may come from strains where their natural genetic background gives only poor or moderate levels of capsular production.

25

As used herein, the term "gene cassette" or simply "cassette", is intended to refer to any DNA segment flanked by one, or both, or part of, the *cps* flanking regions or a *cps* genetic element or DNA sequence which is found between the flanking regions.

30

Prior to the present invention, the foregoing underlying mechanism of genetic recombination of the capsule synthesis genes was unknown, as were the specific sequences involved. A principal contribution of the present inventors is the specific characterization of the

35

- 9 -

individual genes and flanking regions, allowing for their manipulation and individual transfer to hosts. Now, discrete nucleic acid segments, or cassettes, containing a *cps* flanking region, gene or genes, can be readily prepared and easily used in transformation.

#### Hybridization Probes and Primers

Accordingly, in certain embodiments the invention concerns nucleic acid segments that hybridize with *cps* genes and/or flanking regions. The nucleic acid segments will generally be less than about 20 kb in length, and preferably less than about 15 kb in length, or even 10 kb, and will comprise a non-type specific *S. pneumoniae* *cps* gene flanking region, and/or a type-specific *cps* gene, of sufficient length to allow hybridization with a pneumococcal *cps* flanking region and/or gene. Nucleic acid segments that are capable of hybridizing with the 5' flanking region, the 3' flanking region, to both flanking regions, to one or more of the genes designated *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *tnpA* and *'plpA*, and to one or more genes in combination with one or more flanking regions, are encompassed by the invention.

Nucleic acid segments that include a first sequence portion capable of hybridizing to the 5' *cps* gene flanking region and a second sequence portion capable of hybridizing with the 3' *cps* gene flanking region form one aspect of the invention. Such nucleic acid segments may be combined with one or more *cps* genes and may be constructed to form a genetic unit in which the gene (or genes) is located between the two flanking regions. The gene(s) may be from a different *cps* serotype to the flanking regions or they may be from the same *cps* serotype to the flanking regions. Such genetic units may be termed cassettes and may also encompass the form of a circular DNA segment, plasmid, cosmid, or phage. An

- 10 -

isolated fragment of DNA containing DNA such as might be generated by restriction digestion, ligation, and/or PCR methodologies would also be included. The nucleic acid segment, or cassette, may also include other regions of DNA, such as restriction or cloning sites, PCR primers, promoters, antibiotic resistance genes, and the like, as necessary or desired to make a functional genetic unit.

In order to have utility in connection with the present invention, all that is required is that such a nucleic acid segment or genetic unit include a region of sufficient complementarity and size to allow selective cross-hybridization with the target flanking region or gene sequence.

In general, shorter and intermediate length nucleic acid fragments will be useful as hybridization probes and primers, and in particular, for use in PCR, where the primers will allow generation of the entire intervening *cps* sequence.

Thus flanking region and gene fragments on the order of at least about 14-15, 20, 30, 40, 50, or 100 to 200, contiguous complementary nucleotides are contemplated, although sequences of 500, 1,000, or more, nucleotides in length may also be used. The DNA segments may, of course, be of any length within the stated ranges. This is the meaning of "about" in this context, with "about" meaning a range longer or shorter than the stated length, extending to the previously quoted longer and shorter lengths (with about 14 or so still being the minimum length). The ranges thus encompass 1 to 4, 1 to 9, 1 to 49, 1 to 99, and the like, nucleotides in length.

Longer nucleic acid segments and fragments having on the order of up to 1,000, 2,000, 3,000, 5,000, 10,000, 15,000 or longer in length will also have particular



- 11 -

utilities in addition to their function in hybridization embodiments. In particular, longer nucleic acid segments and fragments including selected *cps* coding sequences may be employed in the expression of recombinant proteins, and nucleic acid segments that include one or more *cps* gene sequences positioned between the flanking regions (cassette constructs) may be used in gene transfer embodiments as described above. The DNA segments may, of course, be of any length within the ranges stated above, so long as they function to achieve the desired effect. "About" in this context therefore indicates ranges of from 1 to 999, or 1 to 4,999, and the like, nucleotides longer or shorter than the stated length.

#### 15 Nucleic Acid and Amino Acid Sequences

Exemplary flanking regions sequences, as disclosed herein, are set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:6 (FIG. 7 and FIG. 8).

20 The 5' *cps* gene flanking region is represented by SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4. SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3 corresponds to regions of DNA sequenced in the upstream portion of the 5' flanking region and SEQ ID NO:4 corresponds to the downstream portion of the 5' flanking region and is

25 termed the "repeat" region. DNA between SEQ ID NO: 1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 are also part of the *cps* 5' flanking region. SEQ ID NO:1 is 300 nucleotides in length and corresponds to the last 180

30 nucleotides of *cpsB* and the 5' end of *cpsC* which begins immediately (FIG. 6A and FIG. 7). SEQ ID NO:2 is 261 nucleotides in length and corresponds to a 3' end region of *cpsC* (FIG. 6B and FIG. 7). SEQ ID NO:3 is 262 nucleotides in length and corresponds to part of *cpsE* and

35 the 5' end of the repeat region (FIG. 6C and FIG. 7). SEQ ID NO:4 is 934 nucleotides in length, (nucleotide 1 through 934, FIG. 6D and FIG. 8). The 3' *cps* gene

- 12 -

flanking region is 1307 nucleotides in length, 1 through 1307, SEQ ID NO:6 (nucleotide 5886 through 7192, FIG. 6I and FIG. 8).

5           The inventors have found that certain *S. pneumoniae* nucleotide sequences described in the scientific literature correspond to stretches of sequences from the flanking regions of the present invention. For example, Guidolin et al. (1994) have sequenced 6,322 base pairs of  
10 the 19F *S. pneumoniae* serotype *cps* locus. Sequence analysis indicated that this region contained six complete open reading frames and one partial, which they named *cps19fA* to *cps19fG*. Southern hybridization revealed that *cps19fA* and *cps19fB* were conserved in 12  
15 other *S. pneumoniae* serotypes tested, including serotype 3. *cps19fC* and *cps19fD* also hybridized to Type 3 *S. pneumoniae*. The sequences for *cpsB*, *cpsC* and *cpsE* (SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3), as disclosed herein, are about 99% identical to *cps19fB*, *cps19fC* and  
20 *cps19fD* respectively (Guidolin et al. 1994). However, *cpsE* is truncated at the 3' end with respect to the 19F gene (lacks ~280 nt). The site of the truncation is adjacent to the region designated as the "repeat" sequence (SEQ ID NO:4). Based on blotting data (see  
25 Example 17) part of the repetitive element is in SEQ ID NO:3. This sequence extends 132 nt upstream of the *SacI* site at the start of SEQ ID NO:4, as shown in FIG. 7. Although Guidolin et al. (1994) have sequenced this area in serotype 19f, they do not identify this sequence as  
30 being a gene flanking region and do not suggest its use as part of an *S. pneumoniae* capsular cassette.

Garcia et al. (1993) localized a 781 bp *EcoRV* subfragment of a gene (*cap3-1*) that they proposed was  
35 involved in the transformation to a capsulated phenotype. The first 52 nucleotides of the 781 Garcia et al. sequence correspond to nucleotides 883 to 934 of SEQ ID

WO 95/31548

- 13 -

NO:4. The next 443 nucleotides of the Garcia et al. sequence corresponds to nucleotides 1 to 443 of SEQ ID NO:5. However, the remaining 286 nucleotides of the 781 bp sequence do not show any homology with the instant sequences. As with the above, this article does not identify this sequence as being a gene flanking region and do not suggest its use as part of an *S. pneumoniae* capsular cassette. However, the same group (Arrecubieta et al., 1994) using probes in Southern blotting, identified the upstream region of their cap3A as a sequence common to all pneumoniae strains tested.

The amino acid sequences described by Garcia et al. (1993) include a putative protein of 138 amino acids (CAP3-1), transcribed in the same direction as CpsD, and another truncated open reading frame, transcribed in the opposite direction. The first 117 amino acid residues of CAP3-1 corresponds to the first 117 residues of CpsD (SEQ ID NO:11). CAP3-1 is reportedly similar to the amino-terminus of GDP-mannose dehydrogenase of *Pseudomonas aeruginosa*. The remaining 21 residues of CAP3-1 described by Garcia et al. (1993) have no homology with the remaining 277 amino acid residues of CpsD, suggesting that Garcia et al. had cloned a truncated gene in which the 5' end of the cpsD gene was aberrantly fused to an unidentified DNA sequence.

Pearce et al. (1994) described an *S. pneumoniae* protein and corresponding nucleotide sequence, that they proposed was a permease-like protein involved in peptide transport. They termed this protein PlpA. Nucleotides 484 to 1307 of SEQ ID NO:6 correspond to nucleotides 843 to 1672 of the plpA described by Pearce et al. (1994), and the amino acid SEQ ID NO:15 corresponds to residues 282 to 557 of their PlpA. Previously, Pearce et al. (1993) had described an *S. pneumoniae* protein they called Exp1. The amino acid sequence of Exp1 corresponds to

- 14 -

residues 49 through 209 of SEQ ID NO:15. The inventors have found that the 'plpA located adjacent to the type 3-specific genes lacks about one third of its 5' end when compared to plpA genes located adjacent of the capsule genes of other types such as that used by Pearce et al. (1994). Neither Pearce et al. (1994) nor Pearce et al. (1993) identify this sequence as being involved in capsule synthesis in any way, nor do they suggest that it forms part of a common DNA flanking region.

10

Although certain stretches of nucleotide sequences may have been known in the art, their function, relationship to capsule synthesis and, particularly, their role as interchangeable flanking regions has not previously been described. An important feature of the invention is that the functional characterization of the flanking regions allows, for the first time, for the exchange of *S. pneumoniae* type-specific capsule genes to be manipulated and controlled. This is only possible in light of the inventors discovery of the conserved cps gene flanking regions. Nucleic acid segments, including cassettes and plasmids, that include both 5' flanking region sequences and 3' flanking region sequences are thus one preferred embodiment of the invention. PCR primers that have sequences corresponding to both flanking regions form another preferred embodiment of the invention.

Encoded within the upstream 5' flanking region (SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3) are the partially sequenced genes cpsB, cpsC and cpsE which encode for CpsB (SEQ ID NO:7), CpsC (SEQ ID NO:8 and SEQ ID NO:9) and CpsE (SEQ ID NO:10) (FIG. 6A, FIG. 6B, FIG. 6C and FIG. 7).

35

The invention also includes other cps gene sequences, either alone or in combination with the

- 15 -

flanking region sequences described above. The type-specific portions of the polycistronic *cps* gene locus operon, as disclosed herein, are encompassed within nucleotides 1 through 4951, SEQ ID NO:5 (nucleotides 935 through 5885, FIG. 6E, FIG. 6F, FIG. 6G, FIG. 6H, FIG. 7 and FIG. 8). Beside the open reading frames for the proteins, the sequences also contain putative promoters that direct the transcription of the genes of the *cps* locus. Other promoters, herein termed "recombinant promoters", may also be used to direct the expression of the *cps* genes in accordance with the invention.

The genes encoded within SEQ ID NO:5 and as disclosed herein, include *cpsD*, *cpsS*, *cpsU*, *cpsM* and part of *tnpA*.

*cpsD* is 1277 nucleotides in length, 1 through 1277 of SEQ ID NO:5 (935 through 2211, FIG. 6E and FIG. 8).

*cpsS* is 1267 nucleotides in length, 1277 through 2543 of SEQ ID NO:5 (2211 through 3477, FIG. 6F and FIG. 8).

*cpsU* is 1055 nucleotides in length, 2707 through 3761 of SEQ ID NO:5 (3641 through 4695, FIG. 6G and FIG. 8).

*cpsM* is 1194 nucleotides in length, 3758 through 4951 of SEQ ID NO:5 (4692 through 5885, FIG. 6H and FIG. 8).

*CpsS* is just downstream of *CpsD*, only 15 nucleotides separate a potential start codon for *CpsS* from the stop codon of *CpsD*. Other start codons for *CpsS* are at nucleotides 1311 and 1355 (SEQ ID NO:5). There is a large non-coding region between *cpsS* and *cpsU* (nucleotide 2543 through 2707, SEQ ID NO:5). Where as *cpsD* and *cpsS*

- 16 -

overlap by one nucleotide and *cpsU* and *cpsM* overlap by 4 nucleotides. All four genes are transcribed in the same direction and *cpsD* and *cpsS* are in the same reading frame.

5

Encoded within the 3' flanking region (SEQ ID NO:6) is the truncated sequence for *plpA*, designated '*plpA*', which is 823 nucleotides in length, 484 through 1307 of SEQ ID NO:6 (nucleotides 6370 through 7192, FIG. 6I and FIG. 8). As mentioned above the 5' end is not present in the *plpA* gene of type 3 *S. pneumoniae*. A partial transposase sequence, *tnpA*, is contained between *cpsM* and '*plpA*'. It is transcribed in the opposite orientation to all other genes described herein, and extends from nucleotide 480 through 1, SEQ ID NO:6 to overlap with the *cpsM* gene nucleotide 4951 through 4902, SEQ ID NO:5 a total of 531 nucleotides (nucleotides 5836 through 6366, FIG. 6J and FIG. 8).

20

*cpsD* encodes a 394 amino acid sequence (SEQ ID NO:11) which is homologous to that of the UDP-glucose dehydrogenase (HasB) from *Streptococcus pyogenes*. The deduced amino acid sequence encoded by *cpsS* predicts a protein of 416 residues (SEQ ID NO:12) which is homologous to polysaccharide synthases. *cpsU* encodes a 306 amino acid sequence (SEQ ID NO:13) which is homologous to glucose-1-phosphate uridylyltransferases from several other bacterial species. *cpsM* encodes a 397 amino acid sequence (SEQ ID NO:14) which has homology with phosphoglucomutases from several bacterial species. However, it lacks approximately 25% of the C-terminal present in other phosphomutases and may not encode a functional protein. '*plpA*' encodes a 274 amino acid sequence (SEQ ID NO:15) and *tnpA* encodes a 177 amino acid sequence (SEQ ID NO:16).

35

- 17 -

Thus, in certain particular embodiments the present invention concerns the individual *cps* genes, including segments encoding sequences corresponding to one or more of the *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU* and *cpsM* genes. 5 *tnpA* and *plpA* gene constructs, when combined with other *cps* genes and flanking regions, are also encompassed by the invention. In further embodiments, the invention concerns the respective proteins and polypeptides encoded by the *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *tnpA* and 10 *plpA* genes. The proteins, polypeptides and peptides of the invention may be used in a variety of embodiments, including, e.g., in immunological protocols to generate antibodies that may, in turn, be used in diagnostic embodiments to detect *S. pneumoniae*.

15

It should be noted that in the definition of the genes and proteins, the term "*cps*" is not used to indicate that the gene or protein concerned has a defined role in capsule synthesis in all cases. Rather, it 20 indicates that the gene is located between the *cps* gene flanking regions, i.e., within the *cps* operon, and in close association with other *cps* genes. It should also be noted that, the "*S. pneumoniae* gene region" refers to all genetic elements associated with the *cps* genes, 25 including genes incorporated within the flanking regions. A "genetic element" refers to any DNA that may encode for a protein or polypeptide, regardless of functionality. The utility of the *cps* genes remains that, e.g., they may be used in the same diagnostic manner to identify *S.* 30 *pneumoniae*.

While the present disclosure is exemplified in part through the cloning and sequencing of type 3 *cps* genes, the techniques are equally applicable to the *cps* genes of 35 other capsule serotypes, including any one of the 85 serotypes. For example, using the techniques developed for the characterization of the type 3 *cps* gene region,

- 18 -

the inventors have proceeded to characterize the restriction maps for the type 2 type 6B strains (Example 16, FIG. 11). As expected, the maps of the non-type specific flanking regions were found to be identical from serotype to serotype, whereas the maps for the *cps* gene regions themselves were serotype specific.

#### Diagnostic Embodiments

The *cps* flanking region and gene sequences, and the encoded proteins, may be employed in diagnostic embodiments. For example, the amount of *S. pneumoniae*, or *S. pneumoniae* serotype, present within a biological sample, such as blood, serum or a swab from nose, ear or throat, may be determined by means of a molecular biological assay to determine the level of nucleic acid complementary to the *cps* loci, or even by means of an immunoassay to determine the level of one of the polypeptides encoded by a gene from this locus. The *cps* locus DNA segment used in molecular biological assays may include the non-type specific segments such as the 5' and 3' flanking regions, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6 and any sequence in between, or the region encoding various polypeptides, such as those incorporated within SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, ID NO:5 and SEQ ID NO:6.

In a molecular biological method for detecting *S. pneumoniae*, one would obtain nucleic acids from a suitable sample and analyze the nucleic acids to identify a specific nucleic acid segment complementary to the *cps* loci (whether type- or non-type-specific). The nucleic acid segment will generally be identified by sequence, which method generally includes either; identifying a transcript with a corresponding or complementary sequence, e.g., by Northern or Southern blotting using an appropriate probe or; identifying a transcript with two



- 19 -

or more shorter primers and amplifying with PCR technology.

#### Blotting Techniques

5

To detect *S. pneumoniae*, as may be used to identify a patient with otitis media, pneumococcal pneumonia or pneumococcal meningitis, using a method based upon hybridization and blotting techniques, one may use a  
10 probe with a sequence as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6, including any sequence in between, or an equivalent thereof. This imparts an evident utility to the nucleic acid segments of the present invention.

15

To conduct such a diagnostic method, one would generally obtain sample nucleic acids from the sample and contact the sample nucleic acids with a nucleic acid segment corresponding to the *cps* loci disclosed herein,  
20 under conditions effective to allow hybridization of substantially complementary nucleic acids, and then detect the presence of any hybridized substantially complementary nucleic acid complexes that formed.

25 The presence of a substantially complementary nucleic acid sequence in a sample, or a significantly increased level of such a sequence, in comparison to the levels in a normal or "control" sample, will thus be indicative of a sample that harbors *S. pneumoniae*. Here,  
30 substantially complementary nucleic acid sequences are those that have relatively little sequence divergence and that are capable of hybridizing to the sequences disclosed herein under standard conditions.

35 Where a substantially complementary nucleic acid sequence, or a significantly increased level thereof, is detected in a clinical sample from a patient suspected of

- 20 -

having otitis media, pneumococcal pneumonia or pneumococcal meningitis, for example, this will be indicative of a patient that does have that particular disease. As used herein, the term "increased levels" is used to describe a significant increase in the amount of the *cps* loci nucleic acids detected in a given sample in comparison to that observed in a control sample, e.g., an equivalent sample from a normal healthy subject.

10 A variety of hybridization techniques and systems are known that can be used in connection with the *S. pneumoniae* detection aspects of the invention, including diagnostic assays such as those described in Falkow et al., U.S. Patent 4,358,535.

15 In general, the "detection" of a *cps* locus is accomplished by attaching or incorporating a detectable label into the nucleic acid segment used as a probe and "contacting" a sample with the labeled probe. In such processes, an effective amount of a nucleic acid segment that comprises a detectable label (a probe), is brought into direct juxtaposition with a composition containing target nucleic acids. Hybridized nucleic acid complexes may then be identified by detecting the presence of the label, for example, by detecting a radio, enzymatic, fluorescent, or even chemiluminescent label.

Where one simply desires to distinguish *S. pneumoniae* DNA from the DNA of other bacteria, it is contemplated that the non-type specific region sequences may be employed as probes. However, where one desires to distinguish among different *S. pneumoniae* serotypes, it is contemplated that probes will include both type specific and non-type specific *cps* sequences. The type-specific sequences, being type specific, will selectively hybridize only to corresponding serotypes. Thus one can envision a battery of serotype specific *cps* nucleic acid

- 21 -

hybridization probes can be employed to distinguish and identify serotype DNA samples. In these instances, it is not believed to be necessary to employ restriction enzyme digestion prior to hybridization, but this can be  
5 employed where desired. Alternatively, only one non-type specific sequences may be employed as a "universal probe" that allows detection of restriction fragment length polymorphisms (RFLPs). Typically for RFLP detection, one will employ the more specific hybridization technique of  
10 Southern analysis wherein restriction digestion of the unknown or target DNA is carried out using an enzyme that will cleave either within or surrounding the *cps* gene region (FIG. 4 and FIG. 11) show restriction maps for several of the serotype *cps* gene regions).

15

Many suitable variations of hybridization technology are available for use in the detection of nucleic acids, as will be known to those of skill in the art. These include, for example, *in situ* hybridization, Southern  
20 blotting and Northern blotting. *In situ* hybridization describes the techniques wherein the target nucleic acids contacted with the probe sequences are those located within one or more cells, such as cells within a clinical sample or even cells grown in tissue culture. As is well  
25 known in the art, the cells are prepared for hybridization by fixation, e.g. chemical fixation, and placed in conditions that allow for the hybridization of a detectable probe with nucleic acids located within the fixed cell.

30

Alternatively, target nucleic acids may be separated from a cell or clinical sample prior to contact with a probe. Any of the wide variety of methods for isolating target nucleic acids may be employed, such as cesium  
35 chloride gradient centrifugation, chromatography (e.g., ion, affinity, magnetic), phenol extraction and the like. Most often, the isolated nucleic acids will be separated,

- 22 -

e.g., by size, using electrophoretic separation, followed by immobilization onto a solid matrix, prior to contact with the labelled probe. These prior separation techniques are frequently employed in the art and are generally encompassed by the terms "Southern blotting" and "Northern blotting". Although the execution of various techniques using labeled probes to detect *cps* locus DNA or RNA sequences in clinical samples are well known to those of skill in the art, a particularly preferred method is described in detail herein, in Example 4.

#### PCR Techniques

To detect *S. pneumoniae*, using a method based upon PCR technology of U.S. Patent 4,603,102 (incorporated herein by reference), one may also use one or more probes with a sequence as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 (including sequence in between), SEQ ID NO:5 or SEQ ID NO:6, or an equivalent thereof.

To conduct such a diagnostic method, one would generally obtain sample nucleic acids from a suitable source and contact the sample nucleic acids with two probes or primers corresponding to the *cps* loci disclosed herein, under conditions which allow for hybridization and polymerization to occur. A pair of probes, one corresponding to the 5' flanking region and the other corresponding to the 3' flanking region, would be sufficient to detect the presence of *S. pneumoniae* in a sample and may even be used to indicate the amount of bacteria present. Furthermore the size of the isolated PCR product, when separated by any of the methods as described above, may be sufficient to identify the *S. pneumoniae* serotype. Alternatively the PCR product may be digested with one or more restriction enzymes, which

- 23 -

would enable individual serotypes to be distinguished, using the same principle as RFLP.

In further embodiments, it may be desired to employ other probes corresponding to type specific or non-type specific regions. A battery of serotype specific probes, probes corresponding to type specific DNA regions, may be employed in individual reactions, with a universal probe, a probe corresponding to non-type specific regions. The size of PCR products, with or without prior digestion with restriction enzymes, would distinguish and identify the *S. pneumoniae* serotype.

#### Kits

Kits for use in Southern and Northern blotting or PCR technology, to identify *S. pneumoniae* and/or individuals having, or at risk for developing, otitis media, pneumococcal pneumoniae or pneumococcal meningitis, are also contemplated to fall within the scope of the present invention. Such kits will comprise, in suitable container means, *cps* nucleic acid probes; also, generally, unrelated probes for use as controls; and optionally, one or more restriction enzymes.

#### Characterization of *Streptococcus pneumoniae* Serotypes

In another embodiment of the invention, the non-type specific DNA may be used, to isolate and characterize the type-specific DNA sequence for all or any strain of *S. pneumoniae*. Consequently, the most suitable probes for diagnosing *S. pneumoniae* infection, for use in any molecular biological technique, could be found.

The present invention identifies the common flanking DNA which may be used in hybridizations to target the location of the type-specific *cps* genes for any strain of

- 24 -

*S. pneumoniae*. Once this location has been identified the *cps* genes may then be isolated and characterized by the use of conventional techniques, such as; chromosome crawling, PCR technology, cloning and DNA sequencing, all  
5 are disclosed herein. As mentioned above, this in turn would enable suitable probes from each *S. pneumoniae* strain to be chosen and then used for, diagnostic and research purposes. Furthermore, the genetic elements involved in determining *S. pneumoniae* serotype may be  
10 elucidated, and an understanding of their effect on virulence and evolutionary role may be achieved.

In still more particular embodiments, the invention concerns the preparation and cloning of entire *cps* gene  
15 regions encoding one or more specific *cps* genes of a particular serotype, positioned within a "cassette" for ease of manipulation, e.g. in plasmid preparation or host cell introduction, etc. Thus, *cps* gene cassettes in accordance with the present invention will typically  
20 include left and right hand flanking regions to allow homologous recombination in *S. pneumoniae* host cells.

A preferred method for preparing *cps* gene cassette is through the application of PCR technology wherein left  
25 and right hand primer corresponding to left and right hand flanking region are employed to amplify the *cps* gene coding regions. Of course, the primers employed will typically include at their termini appropriate restriction enzyme site. Thus, the resulting cassette  
30 will preferably include at each terminus a restriction enzyme site of choice. The site, of course, will depend upon the vector that is ultimately employed for manipulation or transformation but may be the specific cleavage site for one or more of the restriction enzymes  
35 shown in FIG. 4 and FIG. 7 or listed in Table 1. The list is, of course, exemplary, and any restriction enzyme

- 25 -

could be employed, as is well known to those of skill in the art.

The present invention also contemplates more  
5 traditional approaches to *cps* gene regions cloning, such  
as by partial fragmentation of *S. pneumoniae* DNA followed  
by cloning into a recombinant cloning host, such as *E.*  
*coli*, and screening by hybridization and antibiotic  
selection (using a selection marker found on the plasmid  
10 or other vector employed for cloning). Of course, if the  
cloning host is not a *S. pneumoniae* host, one may employ  
either type specific or non-type specific *cps* gene  
sequences for screening. In these cases, cassettes that  
are developed may include enzyme restriction sites  
15 naturally found to occur in the flanking regions, such as  
an *SphI* site.

It is contemplated that virtually any type of host  
cell may be employed in connection with the present  
20 invention, depending on the particular application. For  
example, where one simply desires to manipulate *cps* gene  
sequences, any acceptable host may be employed, such as  
an *E. coli*, or even an appropriate eukaryotic host where  
desired. However, where one contemplated producing  
25 capsule polysaccharides, one will desire to employ a gram  
positive host such as bacillus, staphylococcus or  
streptococcal hosts. Particularly preferred will be  
*S. pneumoniae* host, in that it is contemplated that such  
hosts will be more readily amenable to manipulation to  
30 increase capsular polysaccharide production.

The inventors contemplate that a particular  
application, therefore, will be the use of recombinant  
hosts not only for the preparation and manipulation of  
35 *cps* gene sequences, but also in the large scale  
production of capsule saccharides and polysaccharides.  
These polysaccharides are useful for the production of

- 26 -

antigenic haptens and epitopes for use in the production of immunogens. Haptens and epitopes are described herein as portions of a molecule against which an immune response is directed. In conjunction with an molecule  
5 that elicits an immune response, that is an immunogen, an hapten-immunogen complex is able to elicit an immune response.

Generally speaking, where capsule production is  
10 required, one will employ a *S. pneumoniae* host cell into which a selected *cps* gene region is introduced or "cassetted in". First, a DNA segment that includes the selected *S. pneumoniae cps* gene(s) flanked by sufficient  
15 *S. pneumoniae* flanking regions to allow homologous recombination in the *S. pneumoniae* host is identified. It is contemplated that flanking regions on the order of 0.1 to 1 kb will be sufficient to allow recombination to occur. Once an appropriate DNA segment is introduced into the *S. pneumoniae* host, either as genomic DNA or as  
20 a recombinant vector (plasmid), transformed host expressing the introduced *cps* gene are selected.

DNA may be introduced into a suitable host by a variety of mechanisms, including natural transformation  
25 of *S. pneumoniae*, calcium mediated transformation or electroporation of *E. coli*. A particularly preferred method of bacterial transformation includes the steps of making an *S. pneumoniae* competent for transformation by growth in Todd Hewitt broth supplemented with calcium and  
30 bovine serum albumin. Alternatively electroporation may be employed, one makes the bacterial cells, such as the *E. coli* strain LE392, competent in 10% glycerol in water, adds the DNA, and electroporates the cells.

35 Where the *cps* gene DNA segment is introduced in the form of a plasmid, it would be preferable, at certain times, to employ a plasmid that is free of a *S.*



- 27 -

*pneumoniae* origin of replication. Thus, virtually any traditional plasmid (which are designed, e.g. for gram negative hosts such as *E. coli*) may be employed. The reason is that where the plasmid is free of a *S. pneumoniae* origin of replication only those clones that have successfully undergone homologous recombination with the recombinant *cps* gene region will be detected. Stated another way, in this case, there is no requirement of a *S. pneumoniae* origin of replication in order for homologous recombination to occur, and thus homologous recombinants are inherently selected for using such a cloning technique.

Alternatively it may be simpler to introduce the cassette into *S. pneumoniae* on a replicating plasmid with a *S. pneumoniae* origin of replication. In this way, higher levels of polysaccharide production as a result of the elevated copy number of the plasmid (10 to 20) as opposed to the low copy number of the chromosome (1 to 2) would be achieved and homologous recombination need not occur.

Additionally, it is contemplated that there will be some advantage to employing as the starting host a *S. pneumoniae* strain that is a high producer of its own inherent *cps* gene. These high producers will necessarily include the genetic environment to support high production of the newly introduced *cps* complex, and thus will likely be ideal hosts for such production. To achieve such recombinants, all that is required is that the heterologous gene containing the flanking regions, or a genome containing the flanking regions, is introduced into the host cell, and resultant recombinants wherein the homologous gene has been replaced is selected.

35

Although the invention contemplates in particular embodiments the introduction of an entire recombinant *cps*

- 28 -

gene region where capsule saccharide synthesis is desired, the invention also contemplates the introduction of one, two, three or so of the individual *cps* genes. It is contemplated that one or two genes, such as those that control the biosynthesis of sugar precursors may be sufficient to, in and of themselves, confer serotype specific saccharides production on the selected host, or can in of themselves, upregulate capsule production by existing *cps* genes of the host.

10

#### BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

20        **FIG. 1.**    ELISA for the detection of type 3 capsule. Wells of microtiter dishes were coated with crude extracts of capsule material. Type 3 capsule was detected with the monoclonal antibody 16.3 (Briles et al., 1981a). The type 2 strain D39 served as a negative control. Measurements were made in triplicate, and error bars represent standard errors. Values were standardized to protein content. Genotypic designations were based on linkage as determined by transformation mapping.

30        **FIG. 2.**    Insertion-duplication restoration. The cloned fragment and the homologous fragment in the chromosome are represented by the open boxed regions. The dark block represents the mutation in the chromosome of the mutant strain. Insertion of the plasmid clone into the pneumococcal chromosome results in a duplication of the homologous fragment with the plasmid inserted in between. If the recombination occurs to the left of the

35

- 29 -

mutation as shown here, a wild type, full-length copy of the gene is created and function is restored. The plasmid clone leading to restoration spontaneously excises at low frequency through homologous recombination and can therefore be easily recovered by transformation to *E. coli*. pJD330 contains a 2.4 kb *Sau3AI* fragment.

FIG. 3A. Repair of capsule mutations by double crossover recombination event. Mutants were transformed with plasmid subclones of pJD330 and no selection for *Ery*<sup>R</sup> was made. The box in the chromosome represents the region in the mutant chromosome homologous to the fragment cloned in pJD330. The plasmid represents a subclone of pJD330 capable of restoring encapsulation in the mutant strain.

FIG. 3B. Deletion analysis to locate the site of the *cpsA1*, *capD4*, and *Rx1* mutations. The mutations were mapped by transformation with plasmid clones containing the indicated fragments and screening for the mucoid colony phenotype. Identical fragments repaired the *cpsA1*, *capD4*, and *Rx1* mutations. No selection was made for insertion of the plasmids, thus these numbers represent double crossover events. The actual frequencies of repair, shown for the *cpsA1*-containing mutant, are mainly a reflection of the transformation efficiency of the recipient. No encapsulated transformants were obtained when pJY4163 (no insert) was used for transformation. Restriction sites are: M, *MunI*; P, *PstI*; Pv, *PvuII*; R, *RsaI*; S3, *Sau3A I*; Ss, *SspI*; Xb, *XbaI*.

FIG. 4. Physical and genetic map of the type 3 capsule region of *S. pneumoniae* WU2. The restriction map was developed by probing chromosomal digests of WU2 with pJD330 and pJD366 and by probing chromosomal digests of JD770 with pJD330 and pJY4163. The location of primary

- 30 -

clones pJD330 and pJD366 are indicated above the map. Subcloned fragments used to target insertion-duplication mutations are indicated below the map. Insertion of plasmids containing shaded fragments led to loss of capsule production. + or - at the end of a fragment indicates the presence or absence of transcription detected at that point in the chromosome. Insertions of pJD330 and pJD366 are in the orientation to detect transcription to the left. Clones pJD351 and pJD364, which contain the pJD330 and pJD366 fragments, respectively, in the opposite orientation, were also used for the transcription studies. The other plasmids used for insertions were pJD356, pJD337, pJD369, pJD359, pJD362, pJD361, pJD357, and pJD374, in the order shown. The genes indicated by genetic data or suggested by transcription determinations were drawn based on preliminary sequence information. The *cpsDSUM* designations are based on probable functions, as described in the text. Restriction sites are Bg, *Bgl*III; Ev, *EcoRV*; H, *Hind*III; Ha, *Hae*III; M, *Mun*I; P, *Pst*I; Pv, *Pvu*II; R, *Rsa*I; S, *Sac*I; S3, *Sau*3A I; Sa, *Sal*I; Sp, *Sph*I; X, *Xba*I. Restriction sites are not necessarily unique to the entire region.

FIG. 5. Schematic representation of the capsule regions in these strains. Insertions in JD871 and JD872 result from incorporation of pJD366. The insertion in JD803 is pJD330. The shaded square symbol represents type 2 specific DNA; the open square symbol represents type 3 specific DNA; the hatched square symbol represents flanking DNA common to both type 2 and type 3 and; the black square symbol represents pJY4163 or pJY4164. The locations of the probes used are indicated below the map.

FIG. 6A. DNA sequence of the 5' flanking region including partial sequence of *cps3B* and *cps3C* (SEQ ID NO:1) encoding for SEQ ID NO:7 and SEQ ID NO:8.

- 31 -

FIG. 6B. DNA sequence of the 5' flanking region including partial sequence of *cps3C* (SEQ ID NO:2), encoding for SEQ ID NO:9.

5        FIG. 6C. DNA sequence of the 5' flanking region including partial sequence of *cps3E* (SEQ ID NO:3), encoding for SEQ ID NO:10.

10        FIG. 6D. DNA sequence of the "repeat" upstream flanking DNA (SEQ ID NO:4).

15        FIG. 6E. DNA sequence of the region containing *cps3D* (nucleotides 1 through 1277, SEQ ID NO:5). The -35 and -10 hexamers of potential  $\sigma$ -70 type promoters upstream of *cps3D* are underlined and labeled above the sequence. Putative ribosome binding sites are underlined. The precise locations of endpoints of insertion mutations shown in FIG. 7 are indicated by triangles below the sequence and are labeled with the name of the strain containing the given mutation. The locations of spontaneous mutations in *cps3D* are labeled with the sequence of the mutation and the name of the strain containing the mutation. The sequence from the PvuII-SspI fragment of A66R2 began at nucleotide 1921, thus it is possible that additional mutations are present from the PvuII site to this point. Selected restriction sites are shown.

30        FIG. 6F. DNA sequence of the region containing *cps3S* (nucleotides 1277 through 2543, SEQ ID NO:5). The precise locations of endpoints of insertion mutations shown in FIG. 7 are indicated by triangles below the sequence and are labeled with the name of the strain containing the given mutation.

35

FIG. 6G. DNA sequence of the region containing *cps3U* (nucleotides 2707 through 3806, SEQ ID NO:5). The

- 32 -

-35 and -10 hexamers of potential  $\sigma$ -70 type promoters upstream of *cps3U* are underlined and labeled above the sequence. A short region of dyad symmetry just upstream of *cps3U* is overlined. Putative ribosome binding sites are underlined. The precise locations of endpoints of insertion mutations shown in FIG. 7 are indicated by triangles below the sequence and are labeled with the name of the strain containing the given mutation. Selected restriction sites are shown.

10

FIG. 6H. DNA sequence of the region containing *cps3M* (3746 through 4951, SEQ ID NO:5) with corresponding amino acid sequences. Note that the first line of FIG. 6H overlaps the last line of FIG. 6G.

15

FIG. 6I. DNA sequence of the region containing the 3' flanking region including '*plpA*' (SEQ ID NO:6) with corresponding amino acid sequences (SEQ ID NO:15).

20

FIG. 6J. The DNA sequence for a partial transposase A, *tnpA*, located between '*plpA*' and overlapping *cpsM*. The open reading frame is in the opposite orientation starting at nucleotide 6366 through 5836 of FIG. 6I.

25

FIG. 7. Map of the chromosomal region involved in the biosynthesis of *S. pneumoniae* type 3 capsular polysaccharide. Triangles indicate the endpoints of insertion mutations. Filled triangles represent insertions which resulted in loss of capsule production. Open triangles represent mutations which did not result in loss of capsule production. Restriction enzyme sites are: Bg, *Bgl*II; Ev, *Eco*RV; H, *Hind*III; P, *Pst*I; Pv, *Pvu*II; S, *Sac*I; Sa, *Sal*I; Sp, *Sph*I. Also included is a diagram showing the position of SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3, including the corresponding amino acid sequences SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 and SEQ ID NO:10.

30  
35

- 33 -

FIG. 8. Map of the chromosomal region involved in the biosynthesis of *S. pneumoniae* type 3 capsular polysaccharide showing the relationship between SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6, and DNA sequence as described in FIG. 6D through FIG. 6J.

FIG. 9. Location of insertion mutations in the type 3-specific region of the *S. pneumoniae* WU2 chromosome. Schematic illustration of the insertions. The schematic was derived from Southern blot analysis such as that shown in FIG. 11a and FIG. 11b. The ability of the strains to produce type 3 capsule is indicated. Restriction sites are: F, *FspI*; H, *HindIII*; K, *KpnI*; Ms, *MscI*; P, *PstI*; Pv, *PvuII*; X, *XbaI*.

15

FIG. 10. Biosynthetic pathway for type 3 capsular polysaccharide. The functions of the proteins encoded by the type 3-specific genes, based on homologies, genetic, and biochemical data are shown. Additional functions may be necessary for capsule transport or attachment.

FIG. 11. Chromosome maps of the capsule regions in strains of types 2, 3, and 6B. The *SacI-HindIII* fragment (pJD377) from type 3 used for the probe is shown below the maps. Restriction sites are Bg, *BglIII*; F, *FspI*; H, *HindIII*; S, *SalI*; Sac, *SacI*; Sp, *SphI*.

FIG. 12. Production of type 3 capsule. Buoyant densities of parents and derivatives expressing the type 3 capsule. Densities were determined by centrifugation on 0 to 50% Percoll gradients for 30 min at 8,000 x g. Samples were grown in duplicate, and the density of each sample was determined in duplicate gradients. The results shown were obtained with bacteria grown on solid medium. Identical results were obtained from growth in liquid culture.

- 34 -

FIG. 13. Total capsule production. Triplicate cultures of each strain were grown to an  $OD_{600}$  of ~0.5. Duplicate wells of polyvinyl plates were coated with either supernatant fluids or cell sonicates. Total capsule contents of the type 3 parent and the derivatives were determined by using a monoclonal antibody to type 3 capsule. See Table 2, footnote d, for explanation of strain designations.

FIG. 14A. All studies were done in BALB/ByJ mice. See Table 7, footnote d, for explanation of strain designations. Virulence of type 2 derivatives. The parental strains JD770 (3/3) and D39 (2/2) and the derivatives JD803 (2/3) and JD804 (2/3) had similar  $LD_{50}$ s (50 to 75 CFU). For time-to-death studies, groups of mice were infected i.p. with doses approximately 5- to 20-fold above the  $LD_{50}$ . The observed times to death were not related to the dose received. Each circle represents an individual mouse. The median times to death for D39 (2/2) and for derivatives JD803 (2/3) and JD804 (2/3) were 31.5 and 33.0 h, respectively (not significantly different). All three values differed significantly ( $P < 0.005$ ) from that of the type 3 parent JD770 (52h). The values for JD803 and JD804 did not differ from each other.

FIG. 14B. All studies were done in BALB/ByJ mice. See Table 7, footnote d, for explanation of strain designations. Virulence of the type 5 derivatives. Mice were infected i.v. with doses of  $10^3$  to  $10^6$  CFU (10-fold increments) of each type 3 derivative. Doses of  $10^0$  to  $10^2$  and  $10^3$  to  $10^6$  CFU were used from the parental strains DBL5 (5/5) and JD770 (3/3), respectively. The total number of mice used per dose is listed beside each datum point. TK5010\* represents the combined data for TK5010 (5/3), TK5011 (5/3), and TK5012 (5/3). The derivatives did not differ from each other but did differ



- 35 -

significantly from the parental strains JD770 ( $P < 0.0005$ ) and DBL5 ( $P < 0.0001$ ).

FIG. 14C. All studies were done in BALB/ByJ mice.  
5 See Table 7, footnote d, for explanation of strain designations. Virulence of type 6B derivatives. Mice were infected i.p. with doses of  $10^1$  to  $10^6$  CFU of the type 6B derivatives. Doses of  $10^0$  to  $10^3$  CFU and of  $10^3$  to  $10^6$  CFU were examined for the parent strains JD770  
10 (3/3) and DBL1 (6B/6B), respectively. The total number of mice used per dose is listed beside each datum point. TK3028\* represents the combined data for TK3026 (6B/3) and TK3028 (6B/3), which did not differ from each other. However, these strains did differ significantly from the  
15 parental strains JD770 (3/3) ( $P < 0.003$ ) and DBL1 (6B/6B) ( $P < 0.0005$ ).

FIG. 15A. Model for the transfer of type-specific genes. Cassette type-recombination. Replacement of the  
20 recipient's type-specific genes with those of the donor results from homologous recombination between common regions that flank the type-specific cassettes. The open elipsoid symbol represents sequence containing repeated element; the black oblong symbol represents common DNA  
25 upstream of type-specific cassettes and; the open oblong symbol represents common DNA (including *plpA*) downstream of type-specific cassettes.

FIG. 15B. Model for the transfer of type-specific  
30 genes. Binary encapsulation by recombination involving homology at only one end. Integration at one end of the type-specific cassette would occur via homologous recombination through the repeated element. Integration at the other end would result from an apparent  
35 illegitimate recombination. Linkage of the two type-specific cassettes would result if the integration occurred in a repeat element in or closely linked to the

- 36 -

recipient's capsule genes. Symbols represent the same as in FIG. 15A.

FIG. 15C. Model for the transfer of type-specific genes. Binary encapsulation via a transposition-like event. Type-specific cassettes flanked by the repeated element would resolve out of the chromosome and be transferred to recipient cells as circular intermediates. Recombination into the recipient chromosome could occur at a repeat element unlinked (as shown) or linked to the recipient's type-specific genes. Transfer of linear DNA could also yield binary strains as a result of recombination between the two repeat elements that flank the type-specific genes and two repeat elements that are closely linked in the recipient chromosome. Symbols represent the same as in FIG. 15A.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present inventors have determined a genetic and physical map that encompasses the region responsible for the synthesis of the polysaccharide capsule of *S. pneumoniae*. The polysaccharide capsule of *S. pneumoniae* is a potent defense against the immune response of the host organism and is directly involved in bacterial virulence. The capsule locus, *cps*, is basically composed of two functional regions: a central region that contains the genes responsible for capsular biosynthesis and is described herein as the type-specific region, and the non-type specific regions that flank the central biosynthetic type-specific region.

*S. pneumoniae* has evolved a complex 'antigenic shift' mechanism that allows the bacteria to evade the host immune system. The antigenic shift of *S. pneumoniae* occurs via homologous recombination of a type-specific cassette that is replaced through natural transformation.

- 37 -

*S. pneumoniae* is naturally competent allowing for the acquisition of chromosomal DNA from exogenous sources, such as other *S. pneumoniae*. Disclosed herein is evidence identifying the non-type specific regions as  
5 being responsible for providing the sequence identity that allows for homologous recombination cross-over points.

The present inventors have identified and cloned the  
10 region of the *S. pneumoniae* chromosome that contains genes involved in the production of type 3 capsular polysaccharide, and that is specific to type 3 strains. They have also cloned approximately 1-3 kb of DNA flanking both sides of this region and found it to be  
15 common to all capsular serotypes examined. A genetic and physical map of the region is presented in FIG. 6A, FIG. 6B, FIG. 6C, FIG. 6D, FIG. 6E, FIG. 6F, FIG. 6G, FIG. 6H, FIG. 6I, and FIG. 6J. A simplified version of which is shown in FIG. 7 and FIG. 8. The sites of insertion  
20 mutations made within the region are shown in FIG. 7. The regions found by hybridization studies to be specific to type 3 or common to all capsule types are also indicated in FIG. 7. The cloning of the upstream region, creation of insertion mutations, sequence analysis of the  
25 region, hybridization analyses using the upstream region, and an *in vitro* assay of type 3 capsule polymerization are described in the following examples.

The DNA sequence of the region containing the nine  
30 genes; *cps3B*, *cps3C*, *cps3E*, *cps3D*, *cps3S*, *cps3U*, *cps3M*, 'plpA', *tnpA* and the flanking DNA was determined and is presented in FIG. 6A, FIG. 6B, FIG. 6C, FIG. 6D, FIG. 6D, FIG. 6E, FIG. 6F, FIG. 6G, FIG. 6H, FIG. 6I, and FIG. 6J (SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5 and  
35 SEQ ID NO:6) along with the deduced amino acid sequences (SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ

- 38 -

ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16).

Based on genetic, molecular, and biochemical data the inventors have been able to assign putative functions to the type-specific genes in the pathway for type 3 capsular polysaccharide biosynthesis. Two of the genes, *cps3D* and *cps3S*, are required for capsule synthesis. There is substantial evidence to indicate that *cps3D* encodes UDP-glucose dehydrogenase. Described herein is genetic evidence to indicate that several mutations causing the capsule-negative phenotype are located in the gene for UDP-glucose dehydrogenase. The predicted amino acid sequence has characteristics consistent with this function. *Cps3D* shows a high degree of homology to HasB, which is the UDP-glucose dehydrogenase of *S. pyogenes* (Dougherty & van de Rijn, 1993). Within *Cps3D* are sequences homologous to the active site and the NAD-binding site in known UDP-glucose dehydrogenases. It is not possible to perform the standard UDP-glucose dehydrogenase assay on extracts of *S. pneumoniae* due to the presence of a NADH oxidase, which copurifies with the enzyme (Smith, et al., 1960; Smith, et al., 1958). However, extracts from *cps3D* mutants could synthesize type 3 capsule *in vitro* if supplied with UDP-glucuronic acid, i.e., they lacked the ability to convert UDP-glucose to UDP-glucuronic acid and thus lack UDP-glucose dehydrogenase activity.

*Cps3S* is a new member of a family of polysaccharide synthases. All of these polysaccharide synthases for which the structures of the polysaccharides are known produce  $\beta$  (1-4) linked polysaccharides. Thus, it is possible that *Cps3S* forms the  $\beta$  (1-4) linkage in the disaccharide cellobiuronic acid (glcA  $\beta$  (1-4) glc), and that a second enzyme is required to polymerize (i.e., create the  $\beta$  (1-3) linkages) the disaccharides into the

- 39 -

full length polysaccharide. However, HasA, the enzyme most closely related to Cps3S, creates both linkages, a  $\beta$  (1-4) and a  $\beta$  (1-3), in the production of hyaluronic acid capsule (DeAngelis, et al., 1993). HasA has recently  
5 been shown to be sufficient for hyaluronic acid synthesis in heterologous bacteria, given the nucleotide sugar substituents (DeAngelis, et al., 1993a). Because the inventors did not find another required enzyme in the type 3-specific region, Cps3S, like HasA, may synthesize  
10 the polysaccharide by monomer addition.

Neither *cps3U* nor *cps3M* appears to be required for type 3 capsule synthesis. Cps3M and Cps3U should function to convert glucose-6-phosphate into glucose-1-phosphate, and glucose-1-phosphate into UDP-glucose,  
15 respectively (FIG. 10). Since UDP-glucose is necessary for the production of essential cell constituents, including teichoic acid and lipoteichoic acid (Austrian, et al., 1959), the products of other genes may complement the functions lost in Cps3U and Cps3M mutants. There are  
20 at least two plausible reasons for the retention of these genes in the type-specific region. One explanation is that their functions cannot be fully duplicated by the second enzymes. For example, they may play a role in  
25 regulating the amount of polysaccharide produced. Under given conditions, such as during infection, increased production of capsule could be advantageous. The large noncoding region upstream of *cps3U* might be a site at which regulation of *cps3U* and *cps3M* occurs. It should  
30 also be noted that the reactions carried out by Cps3M and Cps3U are each reversible, and the enzymes might be more active in the reverse reaction. Therefore, Cps3U and Cps3M might function to limit the amount of capsule produced.

35

Another possible explanation is that these genes were obtained along with the necessary type-specific

- 40 -

genes in a horizontal transfer from another organism and have not been lost. This theory is consistent with the hybridization data indicating that none of the type-specific genes could be detected even at low stringency  
5 in strains of six other pneumococcal types, including types with related capsular polysaccharide structures (Dillard, et al., 1994). However, if these genes serve no necessary function, it is surprising that they have been maintained in the type 3 cassettes of multiple  
10 strains; i.e., the restriction maps of the type 3 regions of five non-clonal type 3 strains are identical, and all have *cps3U* and *cps3M*.

There are three requirements for a DNA region  
15 to be considered a gene cassette: 1) more than one copy of a gene or set of genes must exist, each specifying the production of a different, but related, product; 2) each copy must be flanked by DNA which is common to all the copies; and 3) cassettes must undergo recombination  
20 resulting in the replacement of one copy by another. More than 80 different capsular serotypes of *S. pneumoniae* have been identified, and the structures of more than half of the polysaccharides have been determined (van Dam, et al., 1990).

25

The presence of multiple types implies that as many different sets of genes exist. The inventors have shown that all the necessary genes specific for the production of capsules of types 2, 3, and 6B (Example 16) are  
30 closely linked to an approximate 1.2 kb fragment present in all capsule types examined. This fragment (corresponding to SEQ ID NO:6 and part of SEQ ID NO:5, see FIG. 4 and Example 5), cloned from the region flanking the type 3-specific genes, contains a gene with  
35 a sequence virtually identical to a gene fragment from type 2 strain, described by Pearce et al., and designated *plpA* (Pearce, et al., 1993). However, the flanking

- 41 -

region from type 3 strain is distinct from the sequence described by Pearce et al. (1993) in that it is missing about one third of the 5' end of the gene, designated 'plpA'. Furthermore Pearce et al. did not identify the location of the plpA gene nor did they attempt to define the sequences on either side of the gene.

The mapping studies reported here confirm that the regions to the right of this fragment are common for at least 4 kb. The regions map differently to the left of the fragment (Example 16), implying that these regions contain the type-specific genes in types 2 and 6B, as shown herein for type 3.

The upstream left flanking region from type 3, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 is common to all capsule types examined (2, 3 and 6B, the Repeat region was also examined in 5, 6A, 8, 9 and 22; Example 17). However, the presence of multiple copies of the Repeat fragment (SEQ ID NO:4) has made the linkage in other types difficult to determine and it is possible that the repeat region may not flank the type-specific genes in other types.

Previous workers have provided biochemical evidence of replacement of capsule gene cassettes. When a type 1 and a type 2 strain were each transformed to type 14 or type 23, they no longer produced UDP-glucuronic acid, implying that the transformants had lost the UDP-glucose dehydrogenase gene (Austrian, et al., 1959). Similarly, a type 1 strain transformed to type 3 encapsulation no longer epimerized UDP-glucuronic acid to UDP-galacturonic acid. Molecular evidence is presented herein for the replacement of the type-specific genes for a type 3 strain transformed to type 2 encapsulation and a type 2 strain transformed to type 3 (Example 6). Together with

- 42 -

Example 18, these observations provide strong evidence for a cassette organization of the type-specific genes.

Since the proposal was put forth that capsule genes  
5 are exchanged through a cassette-type recombination,  
there has always been one glaring exception - binary  
encapsulation. At low frequency, strains of certain  
capsule types transformed with DNA from strains of  
10 certain other capsule types were found to produce both  
polysaccharides (Austrian, et al., 1959). Evidently,  
cassette-type recombination had not occurred in these  
transformants since the genes for the original capsule  
had been maintained. Bernheimer et al., found that  
15 stable binary strains contained the second set of type-  
specific genes at a site unlinked to the recipient's  
type-specific genes (Bernheimer, et al., 1967). Unstable  
binary strains frequently lost the donor type-specific  
genes, which were usually located at a site linked to the  
recipient type-specific genes (Bernheimer, et al., 1967;  
20 Bernheimer and Wermundsen, 1969).

Based on the hybridization data described here  
concerning the flanking regions and replacement of type-  
specific genes, as well as the work of Bernheimer  
25 concerning transformation to binary capsule types  
(Bernheimer, et al., 1968; Bernheimer, et al., 1967;  
Bernheimer, et al., 1969; Bernheimer and Wermundsen,  
1972), the inventors can now propose models for capsule  
type change and binary capsule type formation. Cassette-  
30 type recombination would result from crossover events in  
the homologous flanking regions, leading to replacement  
of the type-specific genes. The left crossover could  
take place in the repeated element in strains containing  
this region linked to the type-specific genes but would  
35 occur in flanking DNA further upstream in strains that  
did not contain the repeat. This type of recombination  
is shown in Fig 19A.



- 43 -

The finding of a repeated element upstream of the type 3 capsule genes (SEQ ID NO:4; Example 17) may provide an explanation for binary encapsulation. It is clear that at least one of the copies of the repeated  
5 fragment in types 2 and 6B is unlinked to the capsule genes, since neither of the type-specific cassettes could be moved with a marker inserted in this location. In type 3, a 2.2 kb *HindIII* fragment containing the repeat element is linked to the type-specific genes but based on  
10 transformation studies, an 8 kb fragment is not. The frequency of binary capsule transformants observed by Bernheimer et al., was significantly lower (10-1000 fold) than related transformations resulting in replacement, leading them to suggest that the recombination event  
15 involved strong homology at only one end. Once integrated at the "atypical" location (unlinked to the type-specific cassette), the genes for the second capsule type could not be moved to the normal location, except by transformation of a strain containing the genes of that  
20 type in the normal location, again suggesting that the non-type specific flanking DNA on at least one end had been lost (Bernheimer, et al., 1967; Bernheimer and Wermundsen, 1972).

25 Since finding that part of the left flanking DNA (SEQ ID NO:4) is repeated in chromosomes of several strains, whereas the right flanking region is only present in one location, it is proposed that the repeated element of the left flanking region may be involved in  
30 the recombination that results in binary capsule type formation. The mechanism proposed by Bernheimer et al., for stable binary strains could involve homologous recombination at a repeated element unlinked to the capsule locus; the recombination at the other end of the  
35 capsule genes would occur by an apparent illegitimate recombination event, as shown in FIG. 15B.

- 44 -

An alternative possibility for the generation of stable binary strains, shown in FIG. 15C, involves a transposition-like event that could result if certain type-specific genes are flanked on both sides by the repeated element. Unstable binary strains could result from either type of integration occurring at repeated elements in, or closely linked to, the recipient's type-specific genes. Instability could result from recombination through genes common to both capsule types, as suggested by Bernheimer et al., for the UDP-glucose dehydrogenases of types 1 and 3. Results presented here provide the basis for examining these possibilities. Binary strains containing the two sets of genes linked are of particular interest since they might recombine to form a novel capsule type. Examination of strains producing related capsule structures may help elucidate the possible mechanisms involved in novel capsule type formation.

Epidemiological studies have indicated that capsule type varies independently of other factors, suggesting that a substantial amount of genetic exchange has occurred (Crain et al., 1990; Coffey et al., 1991; Versalovic et al., 1993). Nonetheless, virulence of clinical isolates appears to correlate with the capsule type expressed (Briles et al., 1992). Taken together, these data suggest that the capsule type has a prominent role in determining virulence. However, epidemiological studies cannot demonstrate a causal relationship between capsule type and virulence due to the variability in the genetic backgrounds of the different serotypes. The characterization of the *S. pneumoniae* capsule locus described here has facilitated the construction of isogenic strains differing only in capsule type. These strains have been used to evaluate the role of capsule type in virulence (Example 18). The cloning of capsule genes and elucidation of the genetic organization of the

- 45 -

capsule locus is a significant step toward understanding antigenic variation and virulence in this pathogen.

#### Cloning of the Type 3 region

5

There are a number of reasons for first cloning the type 3 specific genes: the type 3 capsule has a relatively simple structure that is expected to require a small number of genes for its synthesis; production of type 3 capsule is an easily identifiable phenotype; and finally, the availability of antibodies specific for the type 3 polysaccharide capsule allowed rapid screening for the presence of the capsule, of a large number of isolates, using an ELISA assay as described herein in Example 1. The approach provided and disclosed allowed for the first molecular genetic map of the *cps* gene locus.

The cloning of additional type-specific genes has been accomplished using the information derived from the present invention. Taking advantage of the non-type specific region one can isolate the DNA encoding other type-specific genes by simply obtaining a strain of *S. pneumoniae* known to have a type-specific capsule. Polymerase chain reaction using primers specific for opposite flanking regions and directed toward the opposite flanking non-type specific region are used to amplify the type-specific gene cassette. Where the size of the type-specific region is unknown, restriction fragment length polymorphism analysis, using probes specific for either or both of the non-type specific regions may be used to determine the size.

Antibodies specific for type-specific antigenic epitopes may be used with the present invention to distinguish and evaluate the stability of the *S. pneumoniae* strain prior to, and after cloning of the

- 46 -

region. It will also be used for verifying the directed transfer of type-specific genes to a prospective host. Preferred hosts for polysaccharide capsule production will be gram positive bacteria, in particular members of the *Streptococcus*, *Bacillus* and even *Staphylococcus* species.

#### Selection of Hosts for Type-Specific Capsule Production

10       The present invention provides methods for the selection, isolation and transformation of *Streptococcus* sp. with a type-specific capsule polysaccharide gene locus. A *cps* locus can now be isolated and used to specifically change the capsular phenotype of a selected  
15       host organism. The preferred host organism for use with the present invention is a bacteria that produces high amounts of the capsular polysaccharide.

20       Once a suitable high producing host is identified, it will be used to carry the type-specific genes of choice, as shown in Examples 6 and 18. The organisms can be converted to other serotypes by transforming the high producing recipient bacteria with a gene cassette or with intact genomic DNA. A gene cassette, as previously  
25       mentioned, is a segment of DNA comprising of one or more genes flanked by specific DNA sequences which enables incorporation of the cassette into a recipient's cell chromosome at a specific site or locus via homologous recombination. A cassette may contain type-specific  
30       genes, either alone or in combination with non-type specific genes. Of course, the preferred construct for transfection will be a cassette containing the non-type specific flanking regions.

35       A cassette of the *cps* locus comprising of the *cps* genes and the 5' and 3' flanking regions donated from any one of the 85 *S. pneumoniae* serotypes may be transformed

- 47 -

into a recipient *S. pneumoniae* also belonging to any one of the 85 serotypes. During transformation, recombination would occur in the flanking regions, resulting in the replacement of the recipient's type-specific region by  
5 that of the donor. The capsule type of the recipient would be expected to change to that of the donor.

The introduction of a gene cassette comprising of the *cps* locus or DNA segment or genetic element, into *S.*  
10 *pneumoniae* may be performed by a variety of methods. A particularly preferred embodiment would be to digest the donor *S. pneumoniae* genomic DNA with one or more restriction enzymes such as those described in Table 1, for example, and then separate the entire *cps* locus from  
15 the rest of the genomic DNA by gel purification. This specific DNA segment, *cps* locus or genetic element, may then be ligated into a vector such as a plasmid or cosmid or bacteriophage, and transformed by various methods into the recipient *S. pneumoniae*. Alternatively, the donor *S.*  
20 *pneumoniae*'s entire genomic DNA may be naturally transformed into the recipient by a suitable method, e.g., as described in Example 1. Further still, the donor's genomic DNA may be digested with one or more restriction enzymes and then ligated into a plasmid,  
25 cosmid or bacteriophage, without selecting specifically for the *cps* locus. This may then be transformed into the recipient *S. pneumoniae*.

- 48 -

TABLE 1  
RESTRICTION ENZYMES

	Aat II	GACGT/C
	Acc I	GT/MKAC
5	Acc II	CG/CG
	Acc III	T/CCGGA
	Aci I	CCGC(2/2)
	Acy I	GR/CGYC
	Afl II	C/TTAAG
10	Afl III	A/CRYGT
	Age I	A/CCGGT
	Aha III	TTT/AAA
	Alu I	AG/CT
	AlwN I	CAGNNN/CTG
15	Aoc I	CC/TNAGG
	Apa I	GGGCC/C
	ApaB I	GCANNNNN/TGC
	ApaL I	G/TGCAC
	Asc I	GG/CGCGCC
20	Asu I	G/GNCC
	Asu II	TT/CGAA
	Ava I	C/YCGRG
	Ava II	G/GWCC
	Ava III	ATGCAT
25	Avr III	C/CTAGG
	Bae I	ACNNNNGTAYC
	Bal I	TGG/CCA
	BamH I	G/GATCC
	Bbv I	GCAGC(8/12)
30	Bbv II	GAAGAC(2/6)
	Bcc I	CCATC
	Bcef I	ACGGC(12/13)
	Bcg I	GCANNNNNNCG(12/10)
	Bcl I	T/GATCA
35	Bet I	W/CCGGW

- 49 -

	Bgl I	GCCNNNN/NGGC
	Bgl II	A/GATCT
	Bin I	GGATC (4/5)
	Bpu10 I	CCTNAGC (-5/2)
5	Bpu1102 I	GC/TNAGC
	Bsp1286 I	GDGCH/C
	Bsp106 I	AT/CGAT
	BspC I	CGAT/CG
	BsaA I	YAC/GTR
10	BsaB I	GATNN/NNATC
	BseP I	GCGCGC
	Bsg I	GTGCAG (16/14)
	Bsi I	CTCGTG (5/1)
	BsiY I	CCNNNNN/NNGG
15	Bsm I	GAATGC (1/-1)
	BsmA I	GTCTC (1/5)
	Bsp50 I	CG/CG
	BspG I	CG/CGCTGGAC
	BspH I	T/CATGA
20	BspM I	ACCTGC (4/8)
	BspM II	T/CCGGA
	Bsr I	ACTGG (1/-1)
	BsrB I	GAGCGG (-3/-3)
	BstE II	G/GTNACC
25	BstN I	CC/WGG
	BstX I	CCANNNNN/NTGG
	Cac8 I	GCN/NGC
	Cau II	CC/SGG
	Cfr I	Y/GGCCR
30	Cfr10 I	R/CCGGY
	Cla I	AT/CGAT
	CviJ I	RG/CY
	CviR I	TG/CA
	Dde I	C/TNAG
35	Dpn I	GA/TC

- 50 -

	Dra I	TTT/AAA
	Dra II	TG/GNCCY
	Dra III	CACNNN/GTG
	Drd I	GACNNNN/NGTC
5	Drd II	GAACCA
	Dsa I	C/CRYGG
	Eam1105 I	GACNNN/NGTC
	Eci I	TCCGCC
	Eco3 II	GGTCTC(1/5)
10	Eco47 III	AGC/GCT
	Eco52 I	C/GGCCG
	Eco57 I	CTGAAG(16/14)
	EcoN I	CCTNN/NNNAGG
	EcoR I	G/AATTC
15	EcoR II	/CCWGG
	EcoR V	GAT/ATC
	Esp I	GC/TNAGC
	Esp3 I	CGTCTC(1/5)
	Fau I	CCCGC(4/6)
20	Fin I	GTCCC
	Fnu4H I	GC/NGC
	FnuD II	CG/CG
	Fok I	GGATG(9/13)
	Fse I	GGCCGG/CC
25	Fsi I	R/AATTY
	Gdi II	YGGCCG(-5/-1)
	Gsu I	CTGGAG(16/14)
	Hae I	WGG/CCW
	Hae II	RGCGC/Y
30	Hae III	GG/CC
	Hga I	GACGC(5/10)
	HgiA I	GWGCW/C
	HgaC I	G/GYRCC
	HgiE II	ACCNNNNNNGGT
35	HgiJ II	GRGCY/C



- 51 -

	Hha I	GCG/C
	Hind II	GTY/RAC
	Hind III	A/AGCTT
	Hinf I	G/ANTC
5	Hinl I	GR/CGYC
	Hpa I	GTT/AAC
	Hpa II	C/CGG
	Hph I	GGTCA (8/7)
	Kpn I	GGTAC/C
10	Ksp632 I	CTCTTC (1/4)
	Ksp I	CCGC/GG
	Mae I	C/TAG
	Mae II	A/CGT
	Mae III	/GTNAC
15	Mbo I	/GATC
	Mbo II	GAAGA (8/7)
	Mcr I	CGRY/CG
	Mfe I	C/AATTG
	Mlu I	A/CGCGT
20	Mly I	GACTC (5/5)
	Mme I	TCCRAC (20/18)
	Mnl I	CCTC (7/7)
	Mse I	T/TAA
	Msp I	C/CGG
25	Mst I	TGC/GCA
	Mst II	CC/TNAGG
	Mwo I	GCNNNNN/NNGC
	Nae I	GCC/GGC
	Nar I	GG/CGCC
30	Nci I	CC/SGG
	Nco I	C/CATGG
	Nde I	CA/TATG
	Nhe I	G/CTAGC
	Nla III	CATG/
35	Nla IV	GGN/NCC

- 52 -

	Not I	GC/GGCCGC
	Nru I	TCG/CGA
	Nsi I	ATGCA/T
	Nsp I	RCATG/Y
5	NspB II	CMG/CKG
	Pac I	TTAAT/TAA
	Pal I	GG/CC
	Pfl1108 I	TCGTAG
	PflM I	CCANNNN/NTGG
10	Ple I	GAGTC (4/5)
	PmaC I	CAC/GTG
	Pme I	GTTT/AAAC
	PpuM I	RG/GWCCY
	PshA I	GACNN/NNGTC
15	PspA I	C/CCGGG
	Pst I	CTGCA/G
	Pvu I	CGAT/CG
	Pvu II	CAG/CTG
	RleA I	CCCACA (12/9)
20	Rsa I	GT/AC
	Rsr II	CG/GWCCG
	Sac I	GAGCT/C
	Sac II	CCGC/GG
	Sal I	G/TCGAC
25	Sap I	GCTCTTC (1/4)
	Sau3A I	/GATC
	Sau96 I	G/GNCC
	Sau I	CC/TNAGG
	Sca I	AGT/ACT
30	ScrF I	CC/NGG
	Sdu I	GDGCH/C
	Sec I	C/CNNGG
	SfaN I	GATC/ (5/9)
	Sfc I	CTYRAG
35	Sfe I	C/TYRAG

- 53 -

	Sfi I	GGCCNNNN/NGGC
	SgrA I	CR/CCGGYG
	Sma I	CCC/GGG
	Sna I	CTATAC
5	SnaB I	TAC/GTA
	Spe I	A/CTAGT
	Sph I	GCATG/C
	Spl I	C/GTACG
	Srf I	GCCC/GGGC
10	Sse838 I	CCTGCA/GG
	Ssp I	AAT/ATT
	Stu I	AGG/CCT
	Sty I	C/CWWGG
	Swa I	ATTT/AAAT
15	Taq I	T/CGA
	Taq II	GACCGA (11/9)
	Tfi I	GAWTC
	Tsp45 I	GTSAC
	Tsp E I	AATT
20	Tth111 I	GACN/NNGTC
	Tth111 II	CAARCA (11/9)
	Vsp I	AT/TAAT
	Xba I	T/CTGAGA
	Xcm I	CCANNNNN/NNNNTGG
25	Xho I	C/TCGAG
	Xho II	R/GATCY
	Xma I	C/CCGGG
	Xma III	C/GGCCG
30	Xmn I	GAANN/NNTTC

- 54 -

**Nucleic Acid Hybridization**

The DNA sequences disclosed herein will find utility as probes or primers in nucleic acid hybridization  
5 embodiments. As such, it is contemplated that oligonucleotide fragments corresponding to the sequence(s) of SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3 (including sequences in between), SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6 for stretches of between about 10-14  
10 nucleotides to about 20 or to about 30 nucleotides will find particular utility, with even longer sequences, e.g., 40, 50, 100, 200, 500, and even up to full length, being more preferred for certain embodiments. The ability of such nucleic acid probes to specifically  
15 hybridize to non-type-specific and to type-specific-encoding sequences will enable them to be of use in a variety of embodiments. For example, the probes can be used in a variety of assays for detecting the presence of complementary sequences in a given sample, as may be  
20 used, for example to isolate related type-specific genes. Alternatively, one may use the non-type-specific regions to aid in the isolation and cloning of additional type-specific cassettes. However, other uses are envisioned, including the use of the sequence information for the  
25 preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Nucleic acid molecules having stretches of about 10-14, 20, 30, 50, or even of about 100-200 nucleotides or  
30 so, complementary to SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:3 (including sequences in between), SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6, will have utility as hybridization probes. These probes will be useful in a variety of hybridization embodiments, such as Southern  
35 and Northern blotting in connection with analyzing genomic structure and organization of type-specific genes or both linked and non-linked regulatory genes in diverse

- 55 -

strains of *S. pneumoniae*. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the complementary region may be varied, such as between about 10-14 and about 100 nucleotides, or even up to full length according to the complementary sequences one wishes to detect.

10

The use of a hybridization probe of about 10-14 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 10-14 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 20 nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR technology of U.S. Patent 4,603,102 (incorporated herein by reference) or by introducing selected sequences into recombinant vectors for recombinant production.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of non-type- and of type-specific genes. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. Such hybridization conditions are standard in the art and include low stringency and high stringency. For

- 56 -

applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by 0.02M-0.15M NaCl at temperatures of 50°C to 70°C. One particular example is using the QuickHyb® system (Stratagene's Illuminator™ Nonradioactive Detection System) at 68°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating other genes encoding gene products that are involved in the production of capsule polysaccharides. A preferred embodiment for hybridization conditions is described in detail in Example 4. Further standard hybridization conditions can be found in Sambrook et al., (1989), and are known to those of skill in the art.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate type-specific-encoding sequences from related species, functional equivalents, or the like, less stringent hybridization conditions will typically be needed in order to allow formation of the heteroduplex.

One may also desire to employ other hybridization techniques and to change salt conditions such as varying the amount of salt from between about 0.15M-0.9M. Other parameters that can be modified may be temperature such as those ranging from 20°C to 55°C to optimize the signal-to-noise ratio to reduce unwanted background. The techniques for optimizing hybridization conditions are well known to those of skill in the art and are generally also described within the instruction manual for various reagents and apparatus.

- 57 -

In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as  
5 increased temperature. Thus, hybridization conditions can be readily manipulated as is known to those of skill in the art, and thus will generally be a method of choice depending on the desired results.

10 In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art,  
15 including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or  
20 peroxidase, instead of radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific  
25 hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents  
30 in solution hybridization as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to  
35 specific hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria

- 58 -

required (depending, for example, on the G+C contents, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Following washing of the hybridized surface so as to remove nonspecifically bound probe molecules, specific hybridization is detected, or even quantified, by means of the label.

Longer DNA segments will often find particular utility in the recombinant production of peptides or proteins. DNA segments which encode peptide antigens from about 15 to about 50 amino acids in length, or more preferably, from about 15 to about 30 amino acids in length are contemplated to be particularly useful, as are DNA segments encoding entire cps locus encoded proteins, such as those of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:6. DNA segments encoding peptides will generally have a minimum coding length in the order of about 45 to about 150, or to about 90 nucleotides.

The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, repressors, attenuators, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, nucleic acid fragments may be prepared in accordance with the present invention which are up to about 10,000 base pairs in length, with segments of about 5,000 or 3,000 being preferred and segments of about 1,000 base pairs in length being particularly preferred.



- 59 -

It will be understood that this invention is not limited to the particular nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6, or to the particular amino acid sequences of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16. Therefore, DNA segments prepared in accordance with the present invention may also encode biologically functional equivalent proteins or peptides which have variant amino acids sequences. Such sequences may arise as a consequence of codon redundancy and functional equivalency which are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged.

DNA segments encoding a gene, including the *cpsB*, *cpsC*, *cpsE*, *cspD*, *cspS*, *cspU*, *cspM*, *plpA* and *tnpA* genes may be introduced into recombinant host cells and employed for expressing and producing the type-specific proteins for use in producing type-specific capsule polysaccharides. Alternatively, through the application of genetic engineering techniques, subportions or derivatives of selected type-specific gene locus genes may be employed. Equally, through the application of site-directed mutagenesis techniques, one may re-engineer DNA segments of the present invention to alter the coding sequence, e.g., to introduce improvements to the antigenicity of the protein or to test mutants in order to examine the production of capsule polysaccharides at the molecular level. Where desired, one may also prepare fusion peptides, e.g., where the type-specific coding

- 60 -

regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for immunodetection purposes (e.g., enzyme label coding regions).

5

#### Screening Method Type-Specific Genes

Screening for type-specific genes provides another utility for the *cps* loci of the present invention. A type-specific screening protocol will allow for the epidemiological identification of *S. pneumoniae* and its serotypes at the molecular level. By using one or both of the non-type specific regions as probes one can determine the presence of *S. pneumoniae* from a small sample by immobilizing DNA from the sample onto a solid matrix, for example a slot blot using nitrocellulose, and hybridizing thereto a probe as described in the present invention.

Using either or both of the non-type specific regions of the present invention as a probe or probes one may also screen southern blots. The screening of southern blots may allow one to determine not only the presence of *S. pneumoniae* but also the exact genotype of *S. pneumoniae* present in the sample. In conjunction with densitometric analysis of a southern blot containing multiple serotypes one may determine not only the relative frequency of serotypes within a sample, but in addition one may examine the changing characteristics of the serotypes by examining samples taken at distinct time periods.

It also allows the clinician to determine if a patient is having a recurrence of a particular serotype, if the patient is susceptible to a particular serotype or types, or if a new serotype is increasing in the population.

- 61 -

**Site-Specific Mutagenesis**

Site-specific mutagenesis, also known as site-directed mutagenesis, is a technique useful in the preparation of changes, directed by the laboratory technician, that change the characteristics of genes and their gene products, for the addition of restriction sites, for modifying the activity of promoters, repressors, attenuators, and for directed changes affecting recombination. All of these changes may be produced through specific mutagenesis of the underlying non-type- and type-specific DNA of the present invention. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed

- 62 -

mutagenesis which eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance  
5 herewith is performed by first obtaining a single-  
stranded vector or melting apart the two strands of a  
double stranded vector which includes within its sequence  
a DNA sequence which encodes the type-specific protein or  
proteins encoded by the type-specific gene locus. An  
10 oligonucleotide primer bearing the desired mutated  
sequence is prepared, generally synthetically. This  
primer is then annealed with the single-stranded vector,  
and subjected to DNA polymerizing enzymes such as *E. coli*  
polymerase I Klenow fragment, in order to complete the  
15 synthesis of the mutation-bearing strand. Thus, a  
heteroduplex is formed wherein one strand encodes the  
original non-mutated sequence and the second strand bears  
the desired mutation. This heteroduplex vector is then  
used to transform appropriate cells, such as *E. coli*  
20 cells, and clones are selected which include recombinant  
vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of type-  
specific genes using site-directed mutagenesis is  
25 provided as a means of producing potentially useful  
species, for example a strain having enhanced production  
of type-specific capsular polysaccharides, and is not  
meant to be limiting as there are other ways in which  
sequence variants of other type-specific genes may be  
30 obtained. For example, recombinant vectors encoding  
other type-specific genes, as described herein using the  
non-type specific regions of the capsule polysaccharide  
gene cassette are encompassed.

- 63 -

**Biological Functional Equivalents**

Even though the invention has been described with a certain degree of particularity, it is evident that many alternatives, modifications, and variations will be apparent to those skilled in the art in light of the foregoing disclosure. Accordingly, it is intended that all such alternatives, modifications, and variations which fall within the spirit and the scope of the invention be embraced by the defined claims.

As used in this application, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding the *cps* locus refers to a DNA segment that contains the 5' and/or 3' flanking regions, or the *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *tnpA* or *plpA* coding sequences, yet is isolated away from, or purified free from, total genomic DNA of *S. pneumoniae*. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified 5' or 3' flanking region, or *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *plpA* or even *tnpA* gene, refers to a DNA segment including the coding sequences and, in certain aspects, regulatory sequences, isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this term includes both genomic sequences, cDNA sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides or peptides.

- 64 -

"Isolated substantially away from other coding sequences" means that the locus of interest, in this case the 5' or 3' flanking regions, or *cps B*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *tnpA* or *plpA* coding sequences, forms the significant part of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or cDNA coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the laboratory technician.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that include the 5' or 3' flanking regions denoted by SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4 or SEQ ID NO:6, respectively. DNA segments and vectors that incorporate DNA sequences that encode *CpsB*, *CpsC*, *CpsE*, *CpsD*, *CpsS*, *CpsU*, *CpsM*, *PlpA* or transposase A proteins that include within their amino acid sequences an amino acid sequence as set forth in SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16 respectively are also included.

The term "a sequence as set forth in SEQ ID NO:7-16" means that the sequence substantially corresponds to a portion of SEQ ID NO:7-16 and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of SEQ ID NO:7-16. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein. Accordingly, sequences that have between about 75% and about 85%; or more preferably, between about 86% and about 95%; or even more preferably, between

- 65 -

about 96% and about 99%; of amino acids that are identical or functionally equivalent to the amino acids of SEQ ID NO:7-16 will be sequences that are "essentially as set forth in SEQ ID NO:7-16."

5

Naturally, it will be understood that for the Cps proteins, the definition of "equivalents" in this sense does not extend to distinct, but homologous proteins, such as CpsD and HasB from *Streptococcus pyogenes*; CpsS and HasA from *S. pyogenes*, NodC from *Rhizobium meliloti*; nor CpsU and GtaB from *Bacillus subtilis*. Rather, the scope of equivalents contemplated are such that the changes made still result in a protein that is structurally and functionally a Cps protein.

15

In certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:6, for the flanking regions; and SEQ ID NO:5 for the type-specific encoding regions. The term "as set forth in SEQ ID NO:5" is used in the same sense as described above and means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:5 and has relatively few codons that are not identical, or functionally equivalent, to the codons of SEQ ID NO:5. The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids (as in Table 2).

25  
30

It will be understood that acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the

35

- 66 -

sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to coding nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region, such as promoters.

Allowing for the degeneracy of the genetic code, sequences that have between about 75% and about 85%; or more preferably, between about 85% and about 95%; or even more preferably, between about 95% and about 99% of nucleotides that are identical to the nucleotides of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6 (SEQ ID NO:1-6), will be sequences that are "as set forth in SEQ ID NO:1-6". Sequences that are essentially the same as those set forth in SEQ ID NO:1-6 may also be functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1-6 under standard conditions. Suitable hybridization conditions will be well known to those of skill in the art and are clearly set forth herein, e.g., see Example 4.



- 67 -

Tabl 2. Amino Acids and th Corresponding Codons.

<u>Amino Acids</u>				<u>Codons</u>					
	Alanine	Ala	A	GCA	GCC	GCG	GCU		
	Cysteine	Cys	C	UGC	UGU				
5	Aspartic acid	Asp	D	GAC	GAU				
	Glutamic acid	Glu	E	GAA	GAG				
10	Phenylal- anine	Phe	F	UUC	UUU				
	Glycine	Gly	G	GGA	GGC	GGG	GGU		
	Histidine	His	H	CAC	CAU				
	Isoleucine	Ile	I	AUA	AUC	AUU			
	Lysine	Lys	K	AAA	AAG				
15	Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
	Methionine	Met	M	AUG					
	Asparagine	Asn	N	AAC	AAU				
	Proline	Pro	P	CCA	CCC	CCG	CCU		
	Glutamine	Gln	Q	CAA	CAG				
20	Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
	Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
	Threonine	Thr	T	ACA	ACC	ACG	ACU		
	Valine	Val	V	GUA	GUC	GUG	GUU		
	Tryptophan	Trp	W	UGG					
25	Tyrosine	Tyr	Y	UAC	UAU				

Naturally, the present invention also encompasses DNA segments that are complementary, or essentially complementary, to the sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6. Nucleic acid sequences that are "complementary" are those that are capable of base-pairing according to the standard Watson-Crick complementarity rules. As used herein, the term "complementary sequences" means nucleic acid sequences that are substantially complementary, as may be assessed

- 68 -

by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1-6 under relatively stringent conditions such as those described herein as SEQ ID NO:1, 5 SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6.

The DNA segments of the present invention include those encoding biologically functional equivalent 10 proteins and peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or 15 peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed in the laboratory, may be 20 introduced through the application of site-directed mutagenesis techniques, e.g., to introduce improvements to the antigenicity of the protein or to test *S. pneumoniae* mutants in order to examine capsular productivity at the molecular level.

25 If desired, one may also prepare fusion proteins and peptides, e.g., where the *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *plpA* and *tnpA* coding regions are aligned within the same expression unit with other proteins or 30 peptides having desired functions, such as for purification or immunodetection purposes (e.g., proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

35 As mentioned above, modification and changes may be made in the structure of *CpsB*, *CpsC*, *CpsE*, *CpsD*, *CpsS*, *CpsU*, *CpsM*, *plpA* or *tnpA* and still obtain a molecule

- 69 -

having like or otherwise desirable characteristics. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules, receptors, or catalytic regulation of capsular polysaccharide production. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence (or, of course, its underlying DNA coding sequence) and nevertheless obtain a protein with like (agonistic) properties. Equally, the same considerations may be employed to create a protein or polypeptide with countervailing (e.g., antagonistic) properties. It is thus contemplated by the inventors that various changes may be made in the sequence of SEQ ID NO:7, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13 or SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16 proteins or peptides (or underlying DNA) without appreciable loss of their biological utility or activity.

It is also well understood by the skilled artisan that, inherent in the definition of a biologically functional equivalent protein or peptide, is the concept that there is a limit to the number of changes that may be made within a defined portion of the molecule and still result in a molecule with an acceptable level of equivalent biological activity. Biologically functional equivalent peptides are thus defined herein as those peptides in which certain, not most or all, of the amino acids may be substituted. In particular, the function of given protein must be retained to be an equivalent. Of course, a plurality of distinct proteins/peptides with different substitutions may easily be made and used in accordance with the invention.

- 70 -

It is also well understood that where certain residues are shown to be particularly important to the biological or structural properties of a protein or peptide, e.g., residues in active sites, such residues  
5 may not generally be exchanged. This is the case in the present invention, where CpsD has a putative NAD-binding site and active site region at residues 2 to 29 and 251-263 (SEQ ID NO:11) respectively.

10 Amino acid substitutions are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. An analysis of the size, shape and type of the amino acid side-chain  
15 substituents reveals that arginine, lysine and histidine are all positively charged residues; that alanine, glycine and serine are all a similar size; and that phenylalanine, tryptophan and tyrosine all have a generally similar shape. Therefore, based upon these  
20 considerations, arginine, lysine and histidine; alanine, glycine and serine; and phenylalanine, tryptophan and tyrosine; are defined herein as biologically functional equivalents.

25 In making changes, the hydropathic index of amino acids may be considered. Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics, these are:  
isoleucine (+4.5); valine (+4.2); leucine (+3.8);  
30 phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5);  
35 lysine (-3.9); and arginine (-4.5).

- 71 -

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte & Doolittle, 1982, incorporated herein by reference). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within  $\pm 2$  is preferred, those which are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, i.e. with a biological property of the protein. It is therefore understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0  $\pm 1$ ); glutamate (+3.0  $\pm 1$ ); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5  $\pm 1$ ); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

- 72 -

In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within  $\pm 2$  is preferred, those which are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

#### Antibody Generation

Means for preparing and characterizing antibodies are well known in the art (See, e.g., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988; incorporated herein by reference). This invention thus contemplates the generation of antibodies against the proteins CpsB, CpsC, CpsE, CpsD, CpsS, CpsU, CpsM, PlpA and transposase A or peptides derived therefrom. The CpsB, CpsC, CpsE, CpsD, CpsS, CpsU, CpsM, PlpA and transposase A proteins or peptides may be obtained using standard methods of recombinant expression as is routinely in the art.

The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic composition in accordance with the present invention and collecting antisera from that immunized

- 73 -

animal. A wide range of animal species can be used for the production of antisera. Typically the animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimidobencoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

As is also well known in the art, the immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points

- 74 -

following immunization. A second, booster injection, may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the  
5 immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate MAbs.

MAbs may be readily prepared through use of well-known techniques, such as those exemplified in U.S.  
10 Patent 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, e.g., a purified or partially purified CpsB, CpsC, CpsE, CpsD, CpsS, CpsU, CpsM, PlpA or transposase A protein,  
15 polypeptide or peptide. The immunizing composition is administered in a manner effective to stimulate antibody producing cells. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep frog cells is also possible. The use of rats may provide  
20 certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

25 Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the MAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a  
30 peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of  
35 animals will have been immunized and the spleen of animal with the highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen



- 75 -

with a syringe. Typically, a spleen from an immunized mouse contains approximately  $5 \times 10^7$  to  $2 \times 10^8$  lymphocytes.

5       The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably  
10       are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render them incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

15       Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, pp. 65-66, 1986; Campbell, pp. 75-83, 1984). cites). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14,  
20       FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

25       One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used  
30       is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually  
35       comprise mixing somatic cells with myeloma cells in a 2:1 proportion, though the proportion may vary from about 20:1 to about 1:1, respectively, in the presence of an

- 76 -

agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described by Kohler and Milstein (1975; 1976), and those using polyethylene glycol (PEG), such as 5 37% (v/v) PEG, by Gefter et al. (1977). The use of electrically induced fusion methods is also appropriate (Goding pp. 71-74, 1986).

10 Fusion procedures usually produce viable hybrids at low frequencies, about  $1 \times 10^{-6}$  to  $1 \times 10^{-8}$ . However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in 15 a selective medium. The selective medium is generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and 20 methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine 25 is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are 30 defective in key enzymes of the salvage pathway, e.g., hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B cells can operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the 35 only cells that can survive in the selective media are those hybrids formed from myeloma and B cells.

- 77 -

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide MAbs. The cell lines may be exploited for MAb production in two basic ways. A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide MAbs in high concentration. The individual cell lines could also be cultured *in vitro*, where the MAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations. MAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to

- 78 -

function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

## EXAMPLE 1

Isolation and Characterization of Capsule Mutants

## A. Methods

## 1. Bacterial strains, plasmids, and culture conditions

The bacterial strains and plasmids used are listed herein in Table 3. Culture conditions for *S. pneumoniae* and *E. coli* were previously described (Dillard and Yother, 1991). Erythromycin was used at 0.3  $\mu\text{g/ml}$  and streptomycin was used at 100  $\mu\text{g/ml}$  in *S. pneumoniae* cultures where indicated. Chloramphenicol was used at 1  $\mu\text{g/ml}$  to detect transcription in *S. pneumoniae* isolates carrying pJY4163/4164 chromosomal insertions.

Table 3. Bacterial strains and plasmids.

Strain/ Plasmid	Derivation and properties	Source/Reference
Strain <i>S. pneumoniae</i>		
WU2	Type 3 encapsulated	Briles et al. (1991b)
D39	Type 2 encapsulated	Avery et al. (1944)
Rx1	Non-encapsulated mutant of R36A-A66 hybrid	Ravin (1959) Shoemaker and Guild (1974)
L8-2006	Type 1 encapsulated	Dillard and Yother (1994)
DBL5	Type 5 encapsulated	Yother et al. (1982)

- 79 -

	Strain/ Plasmid	Derivation and properties	Source/Refer nce
	BG9273	Type 6A encapsulated	Dillard and Yother (1994)
	EP2809	Type 8 encapsulated	Dillard and Yother (1994)
	BG5862	Type 9 encapsulated	Dillard and Yother (1994)
	LM100	Type 22 encapsulated	Dillard and Yother (1994)
5	A66R <sub>2</sub>	Non-encapsulated mutant of A66 (Type 3)	Muckerman et al. (1982)
	661	Non-encapsulated mutant of A66 (Type 3)	Bernheimer and Wermundsen (1972)
	JD531	Non-encapsulated mutant of WU2, Em <sup>R</sup> , cpsA1	This work
	JD541	Non-encapsulated mutant of WU2, EM <sup>R</sup> , cpsA2	This work
	JD542	Non-encapsulated mutant of WU2, Em <sup>R</sup> , cpsB1	This work
10	JD551	Non-encapsulated mutant of WU2, Em <sup>R</sup> , cpsB2	This work
	JD571	Non-encapsulated mutant of WU2, Em <sup>R</sup> , cpsB3	This work
	JD600	WU2 Str <sup>R</sup>	This work
	JD611	JD600 x JD531, Em <sup>S</sup> , Str <sup>R</sup> , cpsA1	This work
	JD614	JD600 x JD551, Em <sup>S</sup> , Str <sup>R</sup> , cpsB2	This work
15	JD619	JD600 x JD541, Em <sup>S</sup> , Str <sup>R</sup> , cpsA2	This work
	JD692	JD600 x JD542, Em <sup>S</sup> , Str <sup>R</sup> , cpsB1	This work
	JD816	JD600 x JD671, Em <sup>S</sup> , Str <sup>R</sup> , cpsB3	This work
	JD636	WU2 Str <sup>R</sup> , RM <sup>R</sup> , Nov <sup>R</sup>	This work

- 80 -

	Strain/ Plasmid	Derivation and properties	Source /Reference
	JD752	Isolate of transformation pool W62 x JD811, type 3 encapsulated, Em <sup>R</sup>	This work
	JD770	pJD330 x WU2, type 3 encapsulated, Em <sup>R</sup>	This work
	JD803	JD770 x D39, type 3 encapsulated, Em <sup>R</sup>	This work
	JD871	pJD366 x D39, type 2 encapsulated, Em <sup>R</sup>	This work
5	JD872	JD871 x WU2, type 2 encapsulated, Em <sup>R</sup>	This work
	JD875	pJD366 x DBL5, type 5 encapsulated, Em <sup>R</sup>	This work
	JD908	pJD369 x WU2, non- encapsulated, Em <sup>R</sup>	This work
	<i>E. coli</i> LE392	F <sup>-</sup> hsdR514 (T <sub>k</sub> <sup>-</sup> M <sub>k</sub> <sup>-</sup> ) supE44 supF58 Δ (faciZY) 6 galK2 galT22 met21 trpR55 λ	Tilghman et al. (1977)
10	pJY4163 and pJY4164	Lack origin of replication for <i>S. pneumoniae</i> Promoterless cat gene downstream of multiple cloning site (opposite orientations in 4163, 4164), Em <sup>R</sup>	Yother et al. (1992)
	pJD330	WU2 Sau3A1 fragment cloned into pJY4163 BamHI site, isolated from JD752	This work
	pJD337	pJY4163:: 1.5kb XbaI -PvuII fragment of pJD330	This work
15	pJD343	pJY4164:: 0.4kb MunI fragment of pJD330	This work
	pJD345	pJY4164:: 1.1kb MunI fragment of pJD330	This work
	pJD351	pJY4164:: 2.4kb Sau3AI-Sau3AI fragment of pJD330 orientation opposite pJD330	This work

- 81 -

Strain/ Plasmid	Derivation and properties	Source/Reference
pJD353	pJY4164:: 1.6kb Sau3AI-XbaI fragment of pJD330	This work
pJD357	pJY4164:: 0.3kb MunI- Sau3AI fragment of JD330	This work
pJD359	pJY4164:: 0.6kb PvuII-HaeIII fragment of pJD330	This work
pJD361	pJY4164:: 0.45kb XbaI-PstI fragment of pJD330	This work
5	pJD362	pJY4164:: 0.4kb HaeIII-MunI fragment of pJD330
pJD364	WU2 3.2kb HindIII fragment cloned into pJY4164 HindIII site	This work
pJD366	WU2 3.2kb HindIII fragment cloned into pJY4164 HindIII site, orientation opposite pJD364	This work
pJD368	pJD4164:: 0.45kb RsaI-MunI fragment of pJD330	This work
pJD369	pJY4164:: 0.55 kb PvuII-MunI fragment of pJD330	This work
10	pJD374	WU2 1.2kb Sau3AI fragment cloned in pJY4163
pJD377	pJY4164:: 1.2kb SacI- HindIII fragment of pJD364	This work
pJD380	pJY4164:: 0.36kb Sau3AI-SspI fragment of pJD330	This work

Em<sup>R</sup>, erythromycin resistant; Em<sup>S</sup>, erythromycin sensitive,  
 15 Str<sup>R</sup>, streptomycin resistant; Rif<sup>R</sup>, rifampicin resistant;  
 Nov<sup>R</sup>, novobiocin resistant.

- 82 -

## 2. General DNA techniques

Techniques for DNA fragment isolation, ligations, and plasmid isolation and purification were performed as previously described (Dillard and Yother, 1991; and as  
5 described by Sambrook et al., 1989, the relevant portions incorporated herein by reference). Plasmid screenings were done by scraping colonies from agar plates and incubating these in the lysis solution of Kado and Liu, 3% SDS, 50mM Tris, pH 12.6 (Kado and Liu, 1981). The  
10 lysates were run on agarose gels to determine plasmid sizes.

## 3. Library construction

A plasmid library of random fragments was  
15 constructed by digesting chromosomal DNA from *S. pneumoniae* strain WU2 to completion with *Sau*3AI and ligating these fragments into the *Bam*HI site of pJY4163. The resulting ligation mixture was electroporated into *E. coli* LE392, and transformants were selected on L agar  
20 plates containing 300 µg erythromycin/ml. Individual colonies were patched on erythromycin plates. Each plate contained 100 colonies and constituted a pool. Transformants were pooled by scraping the plates.

## 25 4. Transformations

Encapsulated strains of *S. pneumoniae* were induced to competence as has been described in Yother et al., 1986, incorporated herein by reference. Non-encapsulated strains were made competent for transformation by growth  
30 in Todd Hewitt broth (Difco, Detroit, MI) supplemented with 0.01% CaCl<sub>2</sub>, 0.5% BSA, and 0.5% yeast extract. *S. pneumoniae* cells were allowed to express transforming DNA for 2 h. before plating on agar medium. Electroporation of H<sub>2</sub>O washed *E. coli* LE392 cells  
35 resuspended in 10% glycerol was performed in a BTX Electro Cell Manipulator 600 according to the



- 83 -

instructions of the manufacturer (Biotechnologies and Experimental Research, Inc., San Diego, CA).

#### 5. Preparation of *S. pneumoniae* chromosomal DNA

5 Cultures of *S. pneumoniae* (100 ml) were grown to stationary phase in the presence of 1% choline chloride to prevent autolysis (Briese and Hakenbeck, 1983). The bacteria were centrifuged at 5000 rpm for 10 min and resuspended in 2.5 ml TE buffer (10 mM TrisHCl, 1 mM  
10 EDTA, pH 8.0). SDS was added to 1%, and the cells were lysed at 65°C for 15 min. One fifth volume 5 M KOAc (pH 8) was added, and incubation was continued at 65°C 15 min, followed by incubation on ice for 60 min. Cell debris was removed by centrifugation at 10,000 rpm for 10  
15 min, the supernatant was added to 2 volumes of ethanol, and the DNA was hooked out with a glass rod. The DNA was dried, resuspended in TE, and further purified by CsCl/ethidium bromide buoyant density gradient centrifugation (Radloff et al., 1967).

20

#### 6. Recovery of plasmids resolved from *S. pneumoniae* chromosomes.

A 10 ml culture of late log phase *S. pneumoniae* was centrifuged at 5000 rpm for 10 min. The supernatant was  
25 removed, and the cells were resuspended in 100  $\mu$ l lysis buffer (Saunders and Guild, 1980). Following a 5 min incubation at 37°C, 900  $\mu$ l of Birnboim and Doly solution I was added, and the rest of the alkaline lysis procedure was carried out as for *E. coli* (Birnboim and Doly, 1979).  
30 The resulting preparation contained very little plasmid DNA and was therefore electroporated into *E. coli* where significant quantities of plasmid could be obtained and isolated as described (Birnboim and Doly, 1979).

#### 35 7. Mapping by chromosomal transformation

The integration frequency was used to determine the linkage of spontaneous mutations. Chromosomal DNA from a

streptomycin resistant derivative of the capsule mutant was used to transform the other mutants. The number of encapsulated transformants obtained divided by the number of streptomycin resistant transformants obtained is defined as the integration frequency (Bernheimer and Wermundsen, 1972). Integration frequencies of 0.02 to 0.03 resulted from crosses between mutants carrying mutations in the same locus. Integration frequencies of about 0.3 indicated the mutations were in different loci.

Linkage of plasmid insertions to capsule mutations was determined by using chromosomal DNA from wild type strains carrying non-destructive plasmid insertions to transform capsule mutants. Transformants were selected on erythromycin and screened for encapsulation. The co-transformation frequency determined the degree of linkage.

8. Capsule detection by ELISA

To obtain a crude capsule extract, a 10 ml culture of log phase *S. pneumoniae* cells was centrifuged at 5000 rpm for 20 min, and the supernatant was removed. The pellet was resuspended in 1 ml PBS (50 mM sodium phosphate pH 7.4, 100 mM NaCl). Protein content was determined on 100  $\mu$ l of washed cells using the Bio-Rad kit (Bio-Rad, Richmond, CA). The remaining cells were extracted with an equal volume of 24:1 chloroform/isoamyl alcohol. Following centrifugation, the aqueous phase was precipitated with two volumes of ethanol. The precipitate was resuspended in 1 ml PBS, then treated with RNase at 1 mg/ml and DNase at 0.2 mg/ml for 3 h at 37°C, and pronase at 1 mg/ml for 2 h at 37°C. The extract was then re-extracted and precipitated as before. The precipitate was resuspended in PBS and used to coat the wells of a microtiter plate for ELISA analysis. Values were normalized to protein content. ELISAs were

- 85 -

performed by the standard technique (Ausubel et al., 1987). Monoclonal antibody 16.3 was used to detect type 3 capsular polysaccharide (Briles et al., 1981a). ELISA plates were read at 405 nm in a Biotek model 320 plate reader (Bio-Tek Instruments, Winooski, VT).

#### 9. Percoll gradient centrifugation

For density determinations, 10 ml of log phase cells were centrifuged at 4000 rpm for 10 min, washed once with water, and then resuspended in 1 ml water. A volume of 300  $\mu$ l of cells was loaded on top of a 10 ml 0-100 or 25-100% continuous Percoll gradient. Gradient density marker beads were loaded on top of the gradients. The gradients were centrifuged at 10,000 rpm for 15 min with the brake off. Percoll and density marker beads were purchased from Pharmacia (Piscataway, NJ).

Non-encapsulated strains of *S. pneumoniae* exhibit a higher density in Percoll gradients than encapsulated strains (Briles et al., 1992). Percoll gradients were also used to enrich for encapsulated cells expected to result from low frequency events. Percoll gradients were used to obtain binary capsule type transformants and to enrich for spontaneous revertants to capsule production.

25

#### B. Results

To identify the type 3 capsule region of *S. pneumoniae*, a Sau3AI library of fragments was cloned from the type 3 encapsulated strain WU2 to direct insertions into the chromosome of strain WU2. The library used in the insertion-duplication mutagenesis procedure was constructed by cloning random Sau3AI fragments in the plasmid pJY4163, which carries an erythromycin-resistance marker. Since this plasmid is unable to replicate in *S. pneumoniae*, all erythromycin-resistant transformants should contain insertions at the chromosomal site of the target Sau3AI fragment (Morrison et al., 1984). By

- 86 -

transforming the library of clones into strain WU2, the inventors obtained 5 non-mucoid isolates among 491 erythromycin-resistant transformants. However, further studies involving transformation of the parent strain with either chromosomal DNA from the mutants or plasmids recovered from the mutants showed that the plasmid insertions were neither linked to nor responsible for the capsule mutations.

10 To determine if the non-mucoid isolates were truly deficient in the production of type 3 capsular polysaccharide, several methods were employed. In slide-agglutination assays, none of the five mutants reacted with polyclonal antisera specific for type 3 polysaccharide. Centrifugation through Percoll density gradients revealed that the mutant strains were much denser than the encapsulated parent strain. WU2 cells had a density <1.01 g/ml, whereas all five mutants migrated at 1.09 to 1.10 g/ml. These data suggested that complete capsules were not produced by the mutants. However, these tests might not reveal the presence of short or sparse polysaccharide chains on the cell surfaces or capsular material not translocated to the surfaces. To determine if such material was present, ELISA analysis was carried out on crude cell extracts. Capsular material in the extracts was detected using a monoclonal antibody directed against type 3 polysaccharide. The analysis indicated that mutants JD531 and JD541 (designated mutants A1 and A2 respectively) made no detectable capsular material, whereas mutants JD542, JD551, and JD571 (designated mutants B1, B2 and B3 respectively) made significant levels of reactive material. The common laboratory strain Rx1 was also found to produce significant levels of type 3 capsular material (FIG. 1). Although Rx1 is generally referred to as a non-encapsulated derivative of the type 2 strain D39, it was transformed three times

- 87 -

with chromosomal DNA from derivatives of the type 3 strain A66, chosen twice for type 3 encapsulation, and chosen finally for non-encapsulation (Ravin, 1959; Shoemaker and Guild, 1974).

5

No capsular material could be detected by ELISA in the culture supernatant fluids of mutants JD531 and JD541, indicating that these mutants were not merely defective in attachment of the polysaccharide. Only low  
10 levels of capsular material were detected in supernatants of Rx1, JD542, JD551, and JD571 cultures.

The five mutations resulting in the capsule-deficient phenotype were mapped to two loci by  
15 chromosomal transformation. Reciprocal crosses between the mutants yielded encapsulated transformants for each combination, but not for transformation of a mutant with DNA from the same strain. The mutations were thereby determined to be genotypically distinct. The  
20 transformations also revealed that the mutations in JD531 and JD541 were more closely linked to each other than to the mutations in JD542, JD551, and JD571. Likewise, the mutations in JD542, JD551, and JD571 were more closely linked to one another than to the other two mutations.  
25 The integration frequencies for those mutations judged to be closely linked were similar (0.02 to 0.03) and were ten-fold lower than those judged to be not as closely linked. The genotypic data thus agreed with the phenotypic data, i.e., the two mutations leading to total  
30 loss of capsule synthesis mapped together, and the three mutations causing lack of proper capsular polysaccharide processing mapped together. The loci containing these mutations were temporarily designated *cpsA* and *cpsB*, respectively, and the mutations were named as indicated  
35 in FIG. 1.

- 88 -

Table 4 shows the transformation frequencies of capsule mutations. *S. pneumoniae* strains were transformed with chromosomal DNA from strain JD636, a streptomycin resistant WU2 (Table 3), and streptomycin-resistant transformants were selected. Transformation frequencies are calculated from cultures not induced to competence. With optimal induction, strain WU2 may exhibit a transformation frequency approaching that of the non-encapsulated mutants. However, during the mutagenesis procedure, sub-optimal transformation frequencies were observed (0.0003 to 0.006%). JD908 (Table 3) was also included, it contains an insertion mutation resulting in loss of capsule expression (described in FIG. 4). It would appear that the non-encapsulated mutants are highly transformable, suggesting that the reason for their over-representation in the original transformant population was because of their transformability. The mutagenesis procedure, by selecting for transformability, has enriched for mutants already deficient in capsule (Table 4).

**Table 4. Transformation frequencies of capsule mutants.**

	Recipient	Mutation	Total cfu	Str <sub>R</sub> cfu <sup>b</sup>	Transformation frequency (%)
25	JD531	<i>cpsA1</i>	$2.0 \times 10^5$	$5.0 \times 10^2$	0.3
	JD541	<i>cpsA2</i>	$1.4 \times 10^8$	$7.6 \times 10^2$	0.6
	JD542	<i>cpsB1</i>	$2.4 \times 10^5$	$20 \times 10^2$	0.8
	JD551	<i>cpsB2</i>	$2.0 \times 10^5$	$4.6 \times 10^2$	0.2
	JD571	<i>cpsB3</i>	$2.4 \times 10^8$	$5.0 \times 10^2$	0.2
30	WU2	<i>wt</i>	$0.0 \times 10^5$	1	0.0000
	JD908	<i>cpsS</i>	$10 \times 10^5$	$10 \times 10^2$	0.1

<sup>b</sup> Str<sup>R</sup> streptomycin resistant.

- 89 -

## EXAMPLE 2

Identification of a Clone Containing a Capsule Gene

To identify DNA fragments capable of repairing the  
5 *cpsA1* mutation, JD611, a derivative of JD531 lacking the  
pJY4163 insertion, was transformed with pools of pJY4163  
clones containing *Sau*3AI fragments from strain WU2.  
Transformations and DNA manipulations were performed as  
described in Example 1. In this insertion-duplication  
10 restoration procedure, the plasmid clone is inserted into  
the mutant chromosome, leading to duplication of the  
homologous target fragment and restoration of one wild  
type copy of *cpsA* (FIG. 2). Erythromycin-resistant  
transformants were screened visually for the mucoid  
15 phenotype. One plasmid clone was identified which  
restored encapsulation in the *cpsA1*-containing mutant.  
Due to the duplication of the target fragment, the  
plasmid insertion could resolve out of the chromosome by  
homologous recombination at low frequency. Therefore,  
20 transformation of *E. coli* with DNA from the encapsulated  
transformant and selection for erythromycin-resistance  
allowed recovery of the plasmid, designated pJD330, that  
had repaired the *cpsA1* defect.

25 Transformation of the capsule-deficient mutants with  
pJD330 suggested that the clone contained part of *cpsA*.  
When pJD330 was inserted into the chromosome of the  
*cpsA1*-containing mutant, 56% of the  
erythromycin-resistant transformants became encapsulated  
30 (Table 5). The failure of the remainder of the  
transformants to become encapsulated indicated that the  
cloned fragment contained only one end of the gene. The  
site of the crossover relative to the site of the  
mutation determines whether the mutation will be located  
35 in the incomplete copy of the gene or the full-length  
copy. If the recombination occurs on the left, as shown  
in FIG. 2, the full-length gene is wild type and capsule

- 90 -

is restored. However, if the crossover occurs on the right, the mutation is located in the full-length copy, and no capsule is obtained. This interpretation is supported by the observation that transformants of the *cpsA1* mutant which incorporated the plasmid but did not become encapsulated, spontaneously gave rise to encapsulated, erythromycin-resistant colonies, either by excision and reinsertion of the plasmid or by gene conversion. The *cpsA2* defect was not repaired by pJD330, suggesting that the site of this mutation was either not present on the plasmid clone or was located too near the end of the clone for crossover to repair the defect.

Table 5. Restoration of encapsulation with pJD330.<sup>a</sup>

Recipient	Mutation	Ery <sup>R</sup> cfu <sup>b</sup>	Cps <sup>+</sup> cfu	Cps <sup>+</sup> frequency (%)
JD611	<i>cpsA1</i>	475	267	56
JD619	<i>cpsA2</i>	26	0	0
JD692	<i>cpsB1</i>	13	0	0
JD614	<i>cpsB2</i>	124	0	0
Rx1	<i>cps</i> <sup>-</sup>	158	49	31
661	<i>capD6</i>	56	0	0
A66R2	<i>capD4</i>	4	3	75

<sup>a</sup> Mutants deficient in capsule production were transformed with pJD330 DNA. Transformants were plated on erythromycin to select for those containing chromosomal insertions of pJD330. Erythromycin-resistant transformants were screened for mucoidy. Cps<sup>+</sup> frequency is the ratio of Cps<sup>+</sup>/Ery<sup>R</sup> cfu.

<sup>b</sup> Ery<sup>R</sup>. Erythromycin resistant.

Transformation of pJD330 into the *cpsB*-containing mutants did not restore any of these to encapsulation,



- 91 -

suggesting that *cpsB* is not present on pJD330. However, strain Rx1 was restored to type 3 encapsulation by pJD330 (Table 5).

5

## EXAMPLE 3

*cpsA* Codes for UDP-Glucose Dehydrogenase

Transformation of two previously characterized mutants suggested that pJD330 contains part of the gene for UDP-glucose dehydrogenase. UDPG dehydrogenase is the enzyme which converts UDP-glucose (UDPG) into UDP-glucuronic acid (UDPGA). UDPG and UDPGA are the two nucleotide sugars which are required for type 3 capsule synthesis (Smith et al., 1960). The non-encapsulated mutants A66R<sub>2</sub> and 661 were previously shown to be deficient in the production of UDPG dehydrogenase due to mutations in the locus designated *capD* (Bernheimer and Wermundsen, 1972). Transformation of A66R<sub>2</sub> (*capD4*) with pJD330 restored encapsulation, whereas transformation of 661 (*capD6*) did not (Table 5).

Transformations with chromosomal DNA (Table 6) confirmed that the other *cpsA* and *capD* mutations were closely linked to the region cloned in pJD330. Transformations and other DNA manipulations were performed as described in Example 1. Strain JD770 was obtained by inserting pJD330 into the chromosome of the parental strain WU2. JD770 was found to produce wild type amounts of type 3 capsule. Chromosomal DNA from JD770 was used to transform those mutants which could not be restored to encapsulation by pJD330 (Table 6). The *cpsA2* and UDPG dehydrogenase mutation *capD6* were found to be >90% linked to the plasmid insertion.

- 92 -

Table 6. Restoration of encapsulation with chromosomal DNA linked to pJD330 insertion.<sup>a</sup>

5	Recipient	Mutation	Ery <sup>R</sup> cfu <sup>b</sup>	Cps <sup>+</sup> cfu	Cps <sup>+</sup> frequency (%)
	JD692	<i>cpsB1</i>	17	13	70
	JD614	<i>cpsB2</i>	79	58	73
	JD619	<i>cpsA2</i>	42	39	93
10	661	<i>capD6</i>	6	6	100

<sup>a</sup> Transformants were plated on erythromycin to select for those which had incorporated the region containing the insertion. Erythromycin-resistant transformants were screened for mucoidy. Cps<sup>+</sup> frequency is the ratio of Cps<sup>+</sup>/Ery<sup>R</sup> cfu.

<sup>b</sup> Ery<sup>R</sup> -- erythromycin resistant.

Deletion analysis was performed to more closely localize the sites of the mutations *cpsA1*, *capD4*, and the mutation in Rx1 (*cps*-). By transforming with plasmid subclones and making no selection for insertion of the plasmids, the inventors were able to observe recombination events that occurred as a result of double crossovers between the cloned fragment and its homolog in the chromosome (FIG. 3A). Transformations with several subclones revealed that the sites of the mutations could all be localized to a 250 bp PvuII-SspI fragment common to pJD380 and pJD369 (FIG. 3B). The fact that the same fragment which restores encapsulation in a UDPG dehydrogenase mutant also restores encapsulation in the *cpsA1*-containing mutant suggests that *cpsA* encodes UDPG dehydrogenase. From here on, the *cpsA* loci is designated *cps3D* (see Example 9).

It is known that transformation of UDPG dehydrogenase mutants, including 661, with chromosomal

- 93 -

DNA from a type 1 strain restored type 3 encapsulation by incorporation of the type 1 specific genes at a site other than that occupied by the type 3 genes (Bernheimer et al., 1967, Bernheimer and Wermundsen, 1972). Type 1 capsular polysaccharide contains galacturonic acid; therefore, type 1 strains are expected to produce UDPG dehydrogenase (Austrian et al., 1959). When UDPG dehydrogenase mutants of a type 3 strain were transformed with DNA from type 1 strains, the UDPG dehydrogenase from type 1 complemented the type 3 mutation, allowing the production of both capsular polysaccharides. The UDPG dehydrogenase gene from the type 1 strain was never observed to repair the mutation in the type 3 gene (Bernheimer and Wermundsen, 1969). When the *cpsA1* mutant JD611 was transformed with chromosomal DNA from a type 1 strain, type 3 encapsulated transformants were obtained at a frequency of  $3 \times 10^{-6}$ . This frequency is in agreement with that observed for transformation of mutant 661 (*capD6*) to binary encapsulation (Bernheimer and Wermundsen, 1972) and above the spontaneous reversion frequency ( $< 8 \times 10^{-9}$ ).

#### EXAMPLE 4

##### Genetic and Physical Map of the Type 3 Capsule Region

###### A. Methods

###### 1. Southern Blotting

Southern blotting was performed using the vacuum blotter and chemiluminescent detection system purchased from Stratagene (La Jolla, CA). The PosiBlot<sup>®</sup> 30-30 pressure blotter is part of an integrated system designed to transfer DNA or RNA from agarose gels quickly and efficiently onto solid support matrices, such as Stratagene's hybridization membranes including the Nitrocellulose membranes, Duralose-UV<sup>™</sup> membranes (reinforced nitrocellulose), Duralon-UV<sup>™</sup> membranes (nylon) or Illuminator<sup>™</sup> nylon membranes.

- 94 -

Following electrophoresis the gels are stained in 5  $\mu\text{g/ml}$  of ethidium bromide (EtBr) in water, destained in water and then photographed. Prior to Southern transfer, the gels are pretreated by depurination, denaturation and neutralization. Depurination entails treating the gels with 0.25 N HCl for 5-30 minutes with gentle shaking. Denaturation consist of pouring off the HCl and adding a 0.5 N NaOH and 1.5 M NaCl denaturation solution, enough to cover the gel. The gels are treated for 5 minutes to one hour with gentle shaking. Neutralization involves pouring off the denaturation solution and adding a 0.1 M Tris-HCl (pH 7.5) and 1.5 M NaCl neutralization solution, enough to cover the gel. They are then treated for 5 minutes to one hour with gentle shaking.

15

Gels are then ready for blotting, which is performed with gloved hands. The membrane is prewet in distilled water ( $\text{dH}_2\text{O}$ ) and then in transfer buffer for 5 minutes. 10x SSC buffer-10x SSPE buffer or 25 mM sodium phosphate (pH 6.5) is the transfer buffer for nylon membranes. For nitrocellulose or Duralose-UV membranes, 20x SSC buffer should be used.

20

The membrane and gel are set up in the Posiblot 30-30 pressure blotter and pressure exerted and adjust to 75 mm Hg. Blotting times vary for different gels and depend on the amount and size of the nucleic acid; size, thickness and percentage of gel; and depth of gel wells and volume of sample loaded on the gel; which are routinely optimized. After the allotted blotting time, the position of the wells on the membrane is marked and the gel removed. The gel is generally stained and destained in ethidium bromide to check the efficiency of transfer. The membrane is removed from the device and placed on clean Whatman 3MM paper to allow the excess buffer to be absorbed. Once the membrane is free of standing liquid, but still damp, the membrane and the

30

35

- 95 -

Whatman 3MM paper is placed under a UV light and crosslinked. Alternatively, dry the membrane in a 80°C drying oven for 1-2 hours prior to crosslinking.

5           Boehringer Mannheim's Genius™ Nonradioactive  
Detection System, a chemiluminescence-based, nucleic acid  
detection kit, permits fast, safe and sensitive detection  
of DNA and RNA immobilized on nylon membranes. As little  
10 as 0.1 pg of target plasmid DNA can be detected in a 30-  
minute exposure of the processed blot to X-ray film or,  
in a similar exposure time, 1 pg of a single-copy gene  
can be detected in less than 1.0 µg of genomic DNA. The  
Nonradioactive Detection System can also be used for  
rapid Northern-blot analysis of RNA.

15

After transfer and crosslinking, the membrane is  
prehybridized for 1 hour at 42°C. The labeled probe is  
placed in a microfuge tube containing 100 µl of sonicated  
salmon sperm DNA (10 mg/ml) stock and heated in a boiling  
20 water bath for 5 minutes. This is pulse-spun to collect  
condensation and stored on ice until ready to add to  
hybridization. The probe is added to prehybridization  
solution and hybridized, with shaking, overnight at 42°C  
using standard hybridization solutions as described by  
25 the Genius® protocol. This is washed once for 15 minutes  
at room temperature in 0.1x SSC/0.1% SDS and then washed  
twice for 15 minutes at 60°C in 0.1x SSC/0.1% SDS for  
each wash. The probe is then ready for detection. The  
BRL 1 Kb DNA ladder was used as a molecular weight size  
30 standard (Bethesda Research Laboratories, Gaithersburg,  
MD). Biotin labeled probes for hybridization were  
prepared by nick-translation using the BRL BioNick kit  
(Bethesda Research Laboratories). High stringency  
conditions, as described above should result in the  
35 detection of sequences ≥95% homologous to the probe.  
Reduced stringency was achieved by lowering the wash, or  
hybridization and wash temperatures to room temperature.

- 96 -

At reduced stringency, sequences with 85% homology to the probe should have been detectable.

#### B. Results

5        Using pJD330 as a probe in Southern hybridizations, a physical map of the type 3 capsule region of strain WU2 was developed (FIG. 4). Using the information gained from the chromosomal mapping, the inventors identified and cloned the *Hind*III fragment located to the right of  
10       the pJD330 insert. *Hind*III fragments approximately 3 kb in size were cloned from the WU2 chromosome into pJY4164. By using pJD330 to screen for homology, a clone containing a 3.2 kb insert was identified. This clone, pJD366, was then used to determine the location of the  
15       *cpsB* mutants.

Transformation mapping using JD770 showed that the *cpsB* mutations were about 74% linked to the pJD330 insertion (Table 6). This high frequency indicated that  
20       *cpsB* might be adjacent to the fragment contained in pJD330. When pJD366 was used to transform strains containing the *cpsB* mutations, encapsulation was not restored. Insertion of pJD366 into the WU2 chromosome did not alter the production of type 3 capsule;  
25       therefore, the inventors were able to examine linkage of the insertion to the *cpsB* mutations and determine the relative location of *cpsB*. The pJD366 insertion was found to be only 25% linked to the *cpsB* mutations (as compared to 74% for the pJD330 insertion), suggesting  
30       that *cpsB* is located to the left of the pJD330 insert, as shown in FIG. 4.

In order to localize regions necessary for capsule production, insertion mutations using subclones of pJD330  
35       and pJD366 were made. Transformation of strain WU2 with plasmids containing fragments internal to a gene or operon required for capsule production should result in

- 97 -

loss of encapsulation. Insertion of the plasmid containing the *Sau3AI* - *XbaI* fragment resulted in loss of encapsulation, indicating that this entire 1.6 kb fragment is within a single gene or operon required for capsule synthesis. Likewise, all insertions within this region eliminated capsule production (FIG. 4). Insertion of the plasmid containing the *XbaI*-*PstI* fragment did not disrupt capsule production, indicating that the end of the gene or operon is contained within this fragment. None of the other insertions resulted in loss of capsule, indicating they were not internal to genes or operons required for capsule synthesis.

Since the plasmids used for the chromosomal insertions contain a promoterless *cat* gene, the inventors were able to establish the directions of transcription at the insertion sites. All insertions which contained the *cat* gene in the orientation to detect transcription proceeding to the right (as drawn in FIG. 4) resulted in chloramphenicol resistance. No transcription was detected in the opposite direction (FIG. 4).

#### EXAMPLE 5

##### Homology with Other Capsule Types

If the type-specific genes for capsule production are contained within a cassette, as has been proposed, these genes should show little homology to the type-specific genes from other capsule types. A high degree of homology should exist in the regions flanking the type-specific region (Austrian et al., 1959; Bernheimer and Wermundsen, 1972). The flanking regions may contain common genes necessary for production of capsule of any type or might not be involved in capsule production.

- 98 -

To determine if the regions cloned in pJD330 and pJD366 are specific to type 3 or are present in strains of other capsule types, *Hind*III digested chromosomal DNAs from strains of types 2, 3, 5, 6A, 8, 9, and 22 were Southern blotted and probed with these plasmids or fragments thereof. The fragment contained in pJD330 (the probe used was pJD351, containing 2.4 kb *Sau*3A1 fragment from pJD330) hybridized only with DNA from the type 3 strain, detecting the expected bands at 2.2 and 3.2 kb. No hybridization with the chromosomal DNA of the other six serotypes was detected, nor could the stringency be sufficiently lowered to detect homology in these strains.

When *Hind*III digests of chromosomal DNAs from these same strains were probed with the *Hind*III fragment from pJD366, the expected 3.2 kb band was observed in the type 3 strain, but a 1.1 kb band was found in every other capsule type. Probing with subclones of pJD366 containing the 2.1 kb *Hind*III-*Sac*I fragment or the 1.2 kb *Sac*I-*Hind*III fragment revealed that the homology resided in the more distal 1.2 kb fragment. Therefore, unlike the remaining 4.2 kb of DNA, which could be detected only in the type 3 strains, the 1.2 kb *Sac*I-*Hind*III fragment (pJD377) showed a high degree of homology and could be detected at high stringency in all strains (2, 3, 5, 6A, 8, 9 and 22). This result suggests that this region may be the highly homologous flanking DNA predicted by the model to be adjacent to the type-specific genes.

30

**EXAMPLE 6****Transformation of Capsule Type**

To determine if all the type-specific genes necessary for the production of type 3 capsular polysaccharide were closely linked on the pneumococcal chromosome, strain JD770 was used as a donor in transformation of the type 2 strain D39. Laboratory



- 99 -

techniques were as described in Example 1. Seventy-three erythromycin-resistant transformants were obtained, and all 73 expressed type 3 capsule. No type 2 capsule could be detected by agglutination with type 2 specific  
5 antisera. Using chromosomal DNA from strain JD770, successful transformation of strains of type 5 and type 6B to type 3 encapsulation was also performed (Example 18).

By transforming the type 2 strain D39 with pJD366,  
10 isolates were obtained with the erythromycin-resistance marker closely linked to the type 2 capsule genes. These transformants were the result of recombination between the flanking regions of homologous non-type-specific DNA. Using DNA from one of these isolates, JD871, to transform  
15 the type 3 strain WU2 resulted in 95% co-transformation of type 2 encapsulation with erythromycin resistance. The remaining 5% were found to be type 3 encapsulated, indicating that only the flanking DNA or the plasmid alone was transferred to these isolates. Insertion of  
20 pJD366 into the type 5 strain DBL5 also resulted in a strain - JD875 - with the erythromycin resistance marker linked to the type-specific genes. This strain was successfully used to transform WU2 to type 5 encapsulation.

25

#### EXAMPLE 7

##### Direct Test of the Cassette Model

Transformations and other DNA manipulations were  
30 performed as described in Example 1. Southern Blotting was performed as described in Example 4.

If capsule type change involves a cassette-type recombination mechanism, then transformation of capsule  
35 type should result in replacement of the recipient's type-specific genes by those of the donor. In order to determine if such replacement does occur, DNA was used

- 100 -

from the type 3-specific region to probe *Hind*III digested chromosomal DNA from a strain which was originally type 2 and was transformed to type 3 (JD803), and a strain which was originally type 3 and was transformed to type 2 (JD872) (FIG. 5).

*Hind* III digested chromosomal DNA from strain: 2 (D39 and JD871 from Example 6); 3 (WU2); 3/2 (JD872) and; 2/3 (JD803), were used in Southern blotting. First of all the Southern blot was probed with pJD343 and pJD368. Together these plasmids contain an 800 bp region (*Hae*III-*Mun*I) specific to type 3 and internal to *cpsS* (FIG. 5). The type 3 parent WU2 contained the expected 2.4 kb *Hind*III fragment specific to type 3, whereas neither the type 2 parent D39 nor its derivative JD871, which has pJD366 inserted into the chromosome, contained this fragment. When JD871 (type 2) was used to transform WU2, the resulting strain JD872 was type 2 encapsulated and had lost the 2.4 kb type 3-specific fragment. Similarly, when D39 was transformed with DNA from JD770 (type 3), the resulting strain JD803 was type 3 encapsulated and had gained the type 3-specific fragment.

Reprobing of the same blot with the 1.2 kb *Sac*I-*Hind*III fragment common to all capsule types (pJD377), revealed that the 1.1 kb *Hind*III fragment was present in each of the strains that now produced the type 2 capsule. Further, JD803 had also gained the 3.2 kb *Hind*III fragment present in WU2 (type 3), 2.1 kb of which is type 3-specific. This fragment was also present in JD871 and JD872 since it is contained in the plasmid insert (FIG. 5).

The loss of type 3 genes by the strain converted to type 2 encapsulation indicated that capsule type change

- 101 -

does not occur by addition, but rather by replacement of the type-specific genes.

#### EXAMPLE 8

##### 5     The DNA Sequence and Amino Acid Sequence of cps3D.

#### A. Methods

##### 1. DNA Sequencing

10       Templates for sequencing were prepared from double-stranded plasmid DNA by denaturing with NaOH (2 N) for 5 min at room temperature, and precipitating with 5 M NH<sub>4</sub>OAc and ethanol. DNA was sequenced by the Sanger dideoxy method using the Sequenase 2.0 kit (US  
15     Biochemicals, Cleveland, Ohio) and <sup>35</sup>S-dATP. The oligonucleotide primers 5'-GCCACTATCGACTACGCG-3' (SEQ ID NO:17) and 5'-TCATTTGATATGCCTCCG-3' (SEQ ID NO:18), corresponding to bases 308 to 325 and 445 to 428 of the cloning vectors pJY4163 and pJY4164 (Yother, et al.,  
20     1992), respectively, were used routinely. Primers 5'-GTGAGATAAATAGTAGTGCG-3' (SEQ ID NO:19) and 5'-TCCAGCTCGTGCATAATCT-3' (SEQ ID NO:20), corresponding to bases 3474 to 3493 and 3596 to 3577, respectively, of the type 3 capsule locus (FIG. 6G) were also used. All  
25     oligonucleotide primers were purchased from Oligos, etc. (Wilsonville, OR). DNA sequencing of PCR products was performed using the US Biochem PCR product sequencing kit, according to the directions of the manufacturer. PCR products were sequenced at least twice, from separate  
30     amplification reactions. Greater than 97% of the sequence was obtained for each strand.

##### 2. Sequence analysis

35       The University of Wisconsin Genetics Computer Group programs (Genetics Computer Group, 1991) were used in the analysis of the DNA sequence. Database searches were

- 102 -

performed using the TFASTA program to detect homology with the deduced amino acid sequences. Potential sigma-70 type promoter sequences were located by using the FIND program to search for sequences with six or less mismatches, as compared to the consensus sigma-70 promoter sequence (Mulligan and McClure, 1986), and having a spacing of 15 to 20 bp between the -35 and -10 hexamers. Potential promoter sequences were evaluated using the equations of Harr et al. (Harr, et al., 1983). The sequence presented in FIG. 6D through FIG. 6J (SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6) has been submitted to GenBank for assignment of an accession number.

### 3. Chromosome crawling and inverse PCR

To isolate the 5' end of *cps3D* and upstream DNA, *S. pneumoniae* WU2 chromosomal DNA was first digested with *Ecl136* II (an isoschizomer of *SacI* that results in blunt ends) and separated on a 0.6 % agarose gel. Fragments ranging from 6 to 7 kb were excised and purified using GeneClean (Bio101, La Jolla, CA). A 35 bp *XbaI* UniAmp adaptor (Clontech, Palo Alto, CA) was ligated to the purified fragments. The fragment of interest was then amplified by using a primer for the adaptor and a primer corresponding to the predicted active site sequence (nt 1802 to 1781 of FIG. 6E) of *cps3D*. A 1.8 kb and a 3.8 kb PCR product was obtained. The 1.8 kb PCR product extended from the active site to the *SacI* site upstream of *cps3D* (FIG. 6E). The 3.8 kb PCR product extended from the active site to the second *SacI* site further upstream (FIG. 7). The polymerase chain reaction (PCR) was performed using AmpliTaq DNA polymerase (Perkin-Elmer Corp., Norwalk, CT) in a Perkin Elmer model 480 thermocycler according to the directions of the manufacturer. In a similar manner, the 0.9 kb fragment from the *cps3D* active site to the *EcoRV* site upstream of

- 103 -

*cps3D* was amplified from a 3.5 kb EcoRV fragment from the WU2 chromosome.

To isolate the region 5' of the repeat sequence, a  
5    *SacI*-*MscI* fragment internal to the repeat region  
    (extending from nucleotide 1 to 257 of SEQ ID NO:4) was  
    first cloned into the insertion vector pSF151 (kanamycin  
    resistant, Km<sup>r</sup>), and used to direct an insertion-  
    duplication event into the type 3 *S. Pneumoniae* WU2  
10    chromosome. Chromosomal DNA from the resulting Km<sup>r</sup>  
    strain, JD1008, was digested with *HindIII*, self-ligated,  
    and transformed into the *E. coli*. The resulting Km<sup>r</sup>  
    plasmid, pRS111, contained in the pSF151 vector and DNA  
    flanking the insertion, i.e., DNA extending from the  
15    *HindIII* site in *cps3B* to the *HindIII* site in the repeat  
    sequence (~2.3 kb of *S. pneumoniae* DNA).

#### B. Results

The *cps3D* nucleotide sequence is shown in FIG. 6E  
20    (SEQ ID NO:5). The *Cps3D* amino acid sequence : (SEQ ID  
    NO:11) is highly homologous (56% identity, 73%  
    similarity) to that of the UDP-glucose dehydrogenase  
    (HasB) from *Streptococcus pyogenes* (Dougherty and  
    van de Rijn, 1993). Two other sequences were detected in  
25    the GenBank which shared a high degree of homology with  
    *Cps3D*. These open reading frames from the *Escherichia*  
    *coli* and *Salmonella enteritica rfb* clusters have not been  
    shown biochemically or genetically to be UDP-glucose  
    dehydrogenases (Bastin, et al., 1993), but they share a  
30    high degree of homology with HasB and *Cps3D*.

*Cps3D* (SEQ ID NO:11) has several characteristics  
    consistent with it being UDP-glucose dehydrogenase. The  
    N-terminal amino acid residues 2 to 29 have all the  
35    characteristics of an NAD-binding site (Wierenga, et al.,  
    1986), and this sequence is very homologous to regions  
    from HasB, AlgD (the GDP-mannose dehydrogenase of

- 104 -

*Pseudomonas aeruginosa* [Deretic, et al., 1987]), and the two potential UDP-glucose dehydrogenases from *E. coli* and *S. enteritica*. The homology with AlgD was previously noted by García et al., in the deduced amino acid sequence of the *S. pneumoniae* gene cap3-1 (García, et al., 1993). They suggested that Cap3-1 was the type 3 UDP-glucose dehydrogenase. Sequence ID NO:1 and SEQ ID NO:5 is in complete agreement with that of García et al., from the EcoRV site to the ScaI site (nucleotide 883 to 1377 FIG. 6D and FIG. 6E, containing amino acids 1 to 117, SEQ ID NO:11). However, no other homology was seen, suggesting that these investigators had cloned only the 5' end of the gene.

The Cps3D sequence at amino acid residues 251 to 263 (SEQ ID NO:11) is consistent with this being the active site of the enzyme. This region is identical at the amino acid level with that of HasB and the putative *E. coli* and *S. enteritica* UDP-glucose dehydrogenases. The homology of the active site region of HasB with that of bovine UDP-glucose dehydrogenase and AlgD has been fully described (Dougherty and van de Rijn, 1993). The cysteine at residue 259 (SEQ ID NO:11) of Cps3D contains the essential thiol group of the reactive site (Ridley, et al., 1975). The predicted size of Cps3D (45 kDa) is also similar to the size of the *E. coli* enzyme (47 kDa) (Schiller, et al., 1976).

#### EXAMPLE 9

##### Identification of Capsule Mutants

DNA sequencing and manipulations were performed as described in Example 8.

To determine the nature of the two *cpsA* mutations, identified in Example 1, the regions were amplified from the chromosomes of the mutant strains and sequenced.

- 105 -

Each mutant (JD611 and JD619) contained a single base pair transversion resulting in a premature stop codon in the *cps3D* sequence. The locations of the mutations are indicated in FIG. 6E.

5

To localize the three *cpsB* mutations, also identified in Example 1, located upstream of the UDP-glucose dehydrogenase mutations (*cpsA*), standard PCR or chromosome crawling was used to amplify fragments from the parent type 3 chromosome that contained either the 5' end coding sequence of *cps3D* (nucleotide 1027 to 1802, FIG. 6E), the promoter and the 5' end of *cps3D* (nucleotide 885 to 1802, FIG. 6D and FIG. 6E), or the 5' end of *cps3D* plus approximately 1 kb of upstream DNA (nucleotide 1 to 1802). Each of these fragments was used to transform the capsule-deficient mutants JD614 and JD692. JD692 could be transformed to encapsulation using the 5' end coding sequence of *cps3D*, whereas JD614 was not restored to encapsulation by this fragment but was restored by the fragment containing the 5' end plus 141 bp of upstream DNA, including the promoter. Both of the mutants were restored by the 1.8 kb fragment containing the 5' end of *cps3D* and the upstream DNA, and neither was restored with a fragment containing the 3' end of *cps3D* (nucleotide 1759 to 2385, FIG. 6E). Thus, these upstream mutations are not located in a separate gene but are in either the *cps3D* structural gene or its promoter. Since some capsule material is produced by these mutants, a mutation within the coding region (as in JD692) must be a missense mutation or an in-frame deletion or insertion which reduces the activity of the enzyme. The mutation in JD614 may be in the promoter, and thus, a promoter down mutation, or it may be in the structural gene but too close to the beginning of the gene for recombination and repair to occur with the fragment used.

10

15

20

25

30

35

- 106 -

Amplification and sequencing of the 250 bp PvuII-SspI fragment from the mutant strains A66R2 and Rx1 showed that each contained a missense mutation in the *cps3D* coding sequence (FIG. 6E).

5

**EXAMPLE 10****DNA Sequences of *cps3S* and *cps3U***

DNA sequencing and analysis was performed as described in Example 8.

The region just downstream of *cps3D* contains a second gene, *cps3S*, that is required for type 3 capsular polysaccharide biosynthesis. An open-reading frame, 1248 bp in length, is transcribed in the same direction as *cps3D* and is in the same reading frame (SEQ ID NO:5). The direction of transcription is in agreement with that determined using cat insertions as described in Example 4. Only 15 bp separate a potential start codon for *cps3S* from the stop codon of *cps3D*. The sequence AGGGG just upstream of the putative start codon may serve as a ribosome binding site (FIG. 6E), or due to the close proximity of *cps3D*, no ribosome binding site may be necessary. The deduced amino acid sequence of Cps3S predicts a protein of 48 kDa (SEQ ID NO:12), if the first start codon at nucleotide 1 (nucleotide 2227 in FIG. 6F) is utilized. Other potential start codons are located at nucleotide 1 plus 19 and +61 (nucleotide 2245 and 2287 FIG. 6F, respectively), however neither of these are positioned near a ribosome binding site.

A short region of dyad symmetry was detected downstream of *cps3S* at nucleotide 3718 to 3738 (FIG. 6F). The scores for primary and secondary structure yielded by the TERMINATOR program of the GCG sequence analysis package (p=3.95, s=22) suggest that this region could function as a weak rho-independent terminator. However,



- 107 -

this sequence is 241 bases past the *cps3S* stop codon and closer to the start of the next open reading frame, i.e., *cps3U*. Therefore, the sequence may have more to do with the expression of *cps3U* than of *cps3S*. In fact, a potential promoter sequence for *cps3U* was detected upstream of the region of dyad symmetry, suggesting the potential structure could serve as an attenuator of *cps3U* expression. The *cpsU* open reading frame (SEQ ID NO:5), 918 bp in length, is transcribed in the same direction as *cps3D* and *cps3S*, and is predicted to encode a protein of 34 kDa (SEQ ID NO:13).

#### EXAMPLE 11

##### Cps3S is Homologous to Polysaccharide Synthases

A search of the GenBank revealed that the predicted Cps3S protein (SEQ ID NO:12) is homologous to polysaccharide synthases. The greatest degree of homology was found with HasA, the hyaluronic acid synthase from *S. pyogenes* (23% identity, 50% similarity) (DeAngelis, et al., 1993b; Dougherty and van de Rijn, 1994). Hyaluronic acid consists of alternating N-acetyl glucosamine and glucuronic acid residues. Hyaluronic acid and the pneumococcal type 3 capsule are similar in structure in that both are composed of  $\beta(1-4)$  linked repeating disaccharide units containing glucuronic acid. Like pneumococcal type 3 capsule, hyaluronic acid capsule contains both  $\beta(1-3)$  and  $\beta(1-4)$  linkages, however the linkage to glucuronic acid is  $\beta(1-4)$  in hyaluronic acid but  $\beta(1-3)$  in type 3 capsule (Reeves and Goebel, 1941). Homology was also seen between Cps3S and NodC from *Rhizobium meliloti* (21% identity, 47% similarity). NodC is necessary for the synthesis of nodulation factor, a substituted oligosaccharide consisting of  $\beta(1-4)$  linked N-acetyl glucosamine residues (Lerouge, et al., 1990). It has previously been noted that HasA and NodC are homologous to polysaccharide synthases, including FBF15

- 108 -

of *Stigmatella aurantiaca*, pDG42 of *Xenopus laevis*, and chitin synthases from both *Saccharomyces cerevisiae* and *Candida albicans* (DeAngelis, et al., 1993b; Dougherty and van de Rijn, 1994; Atkinson and Long, 1992; Debellé, et al., 1992). Cps3S is also homologous to these proteins. These results suggest that Cps3S is the type 3 capsular polysaccharide synthase.

The PILEUP program was used to align the amino acid sequences of the bacterial polysaccharide synthases Cps3S (SEQ ID NO:12), HasA, NodC, and FBF15. Only a few clusters of amino acids are found to be conserved in all four proteins. A few of these, GKR (residues 131 to 133 SEQ ID NO:12), an acidic region VDSG (153 to 156 SEQ ID NO:12), DRXLT (256 to 260 SEQ ID NO:12), QQXRW (292 to 296 SEQ ID NO:12), and WXTR (418 to 421 SEQ ID NO:12), are also found in the eukaryotic polysaccharide synthases.

Since all four proteins contain highly hydrophobic stretches, hydrophobic amino acids are found conserved at several locations throughout the proteins. Four hydrophobic stretches identified in Cps3S are found in all four proteins. These regions may span the cell membrane. This hypothesis has been supported for NodC. Immunogold labeling revealed a surface location for NodC, and the C-terminal hydrophobic region was shown to direct the insertion of an alkaline phosphatase fusion protein to the cell membrane (Johnson, et al., 1989; John, et al., 1988). Earlier studies indicated that the type 3 capsule synthesizing activity also has a membrane location (Smith, et al., 1961). The last hydrophobic stretch may be required for the function of Cps3S since the insertion in JD897 which eliminated this region (the last 45 amino acids of the protein) resulted in loss of capsule production (FIG. 6F). Expression of Cps3S in *E. coli* was, like that of NodC, lethal to the host.

- 109 -

**A. Method****1. Expression of *Cps3S***

A 2.1 kb *Sau3AI*-*PstI* fragment containing the 3' end  
5 of *cps3D* and the entire *cps3S* gene was cloned from  
pJD351 into the expression vector pKK223-3 (Brosius and  
Holy, 1984) at the polylinker *BamHI*-*PstI* sites to yield  
pJD424. Cultures of *E. coli* TG-1 (Sambrook, et al.,  
1989) or TG-1 transformants were grown to exponential  
10 phase, at which time isopropyl-b-D-thiogalactoside (IPTG)  
was added to a concentration of 1 mM to induce expression  
from the *tac* promoter of pKK223-3.

Transformations and other DNA manipulations were  
15 performed as described in Example 1.

**B. Results**

The sequence from residues 211 to 233 in NodC was  
noted for the large number of cysteine residues. It has  
20 been suggested that this region participates in the  
binding of divalent cations which are necessary for the  
production of chitin and chitin-like molecules (Atkinson  
and Long, 1992). Type 3 capsule synthesis requires  $Mg^{++}$   
(Smith, et al., 1960). Although this region in *Cps3S*  
25 contains only one cysteine, the region is highly  
conserved between all four proteins.

The GenBank search also revealed that *Cps3S* has  
homology over short stretches to the rhamnosyl  
30 transferase RfbN from *Salmonella enteritica*, which is  
necessary for the production of O-antigen in type B  
strains. This enzyme creates an  $\alpha(1-4)$  linkage to  
mannose in the O-antigen repeat unit. The homologous  
regions are a subset of those conserved regions common to  
35 HasA, NodC, and *Cps3S*, but the best homology is seen in  
the region 229 to 278 (SEQ ID NO:12).

- 110 -

In the production of Group B type III capsular polysaccharide, the galactosyl transferase CpsD transfers a galactose to a molecule located in the cell membrane. Rubens et al. (1993), suggested that the acceptor may be dolichol or a related molecule, and identified a region of CpsD with homology to putative dolichol binding regions of several proteins. Although it is not clear that such sequences are actually involved in dolichol recognition or binding (Schutzbach, et al., 1993), several similar regions (e.g., at residues 7 to 20, 21 to 38, and 388 to 401, as numbered in SEQ ID NO:12) are present in CpsS. Since the putative dolichol binding motif [FL(F/I)VXFXXI(P/L)FXFY] (Albright, et al., 1989; Kelleher, et al., 1992) is a highly hydrophobic sequence that is rich in phenylalanines, the sites in Cps3S may actually reflect the hydrophobicity of the molecule and the A-T rich bias in the DNA sequence rather than indicating a specificity for dolichol-like molecules. It is not known whether *S. pneumoniae* utilizes an intermediate acceptor in capsule synthesis, however the capsular polysaccharides of several serotypes have been found to be covalently linked to the cell wall. Type 3 capsule, by contrast, is not covalently linked to the cell and is generally considered an exopolysaccharide (Sorensen, et al., 1990). Therefore, if Cps3S does use a membrane bound acceptor, it is likely not the final acceptor.

#### EXAMPLE 12

Cps3U is Homologous to Glucose-1-Phosphate Uridyltransferases and Cps3M is Homologous to Phosphomutases

The gene downstream of *cps3S* is designated as *cps3U* (SEQ ID NO:5) based on its probable function. The amino acid sequence of Cps3U (SEQ ID NO:13) showed a high degree of homology with glucose-1-phosphate

- 111 -

uridylyltransferases from several other bacterial species. The highest degree of homology was found with GtaB from *Bacillus subtilis* (55% identity, 73% similarity). The active site of glucose-1-phosphate uridylyltransferase has not been characterized from any of the bacterial enzymes, however, the active site in the enzyme from potato tuber (*Solanum tuberosum*) has been investigated. Kazuta et al., recognized 5 lysine residues present at the active site (Kazuta, et al., 1991), and by mutational studies Katsube et al., showed that one of these residues was important for function, and a second was absolutely required (Katsube, et al., 1991). Cps3U contains 24 lysines, six of which are absolutely conserved among the six bacterial glucose-1-phosphate uridylyltransferases in the database. Only one region from Cps3U containing a conserved lysine can be aligned well with the potato tuber enzyme sequence. It is homologous to the region containing the required lysine.

20

The final gene in SEQ ID NO:5, is *cpsM* with a deduced amino acid sequence (SEQ ID NO:14). The CpsM amino acid sequence revealed significant homology to both phosphoglucomutases (PGM) and phosphomannomutases (PMM) from a diverse group of microorganisms. Contained with CpsM is a phosphoserine signature sequence (GIMVTASHTPAPFNG) conserved within the reported active sites of both PGMs and PMMs. However, approximately 15% of the C terminus present in other phosphomutases, and apparently more important for their function, is absent from CpsM. Phosphomutase activity from a recombinant CpsM was not detected in *E. coli*, suggesting that *cpsM* may encode a non-functional protein.

30

- 112 -

## EXAMPLE 13

*cps3S* and *cps3D* are Transcribed as an Operon

## A. Methods

5        Southern blotting was performed as described in  
Example 4, all other DNA manipulations, including  
insertion deletion mutations, were performed as described  
in Example 1. The locations of mutations can be seen in  
FIG. 6E and FIG. 6F.

10

## B. Results

Use of fragments subcloned from the *cps3DSU* region  
to direct insertion-duplication mutations in the parent  
type 3 chromosome resulted in several mutants that  
15        produced no detectable capsule (FIG. 9) and exhibited the  
extremely rough phenotype described by Taylor (1949).  
The colonies were very small and rough, and the cells  
clumped when grown in liquid culture. DNA sequencing  
revealed that the sites of the mutations are within *cps3S*  
20        (FIG. 6F). The lack of capsule production in these  
mutants must be due to loss of *cps3S* expression, rather  
than to a polar effect on downstream genes, since  
insertions within *cps3U* or *cps3M*, the next genes  
downstream, had no apparent effect on capsule production.

25

Molecular and genetic evidence suggest *cps3S* is in  
an operon with *cps3D*. Sequence analysis revealed no  
potential promoter sequences in the region upstream of  
*cps3S* (FIG. 6E, SEQ ID NO:5). The phenotypes of several  
30        insertion mutants also suggest that no promoter is  
located in the 3' end of *cps3D* and that *cps3S* is  
transcribed from the *cps3D* promoter. The sites of these  
insertions are shown in FIG. 6E, FIG. 6F and FIG. 7,  
however, the structures of the mutations are more fully  
35        illustrated in FIG. 9. To insure that the plasmids had  
inserted as expected for insertion-duplication mutations,

- 113 -

chromosomal DNA from the mutant strains was subjected to Southern blot analysis.

Insertion mutants were digested with *MscI*/*FspI* for JD982, *MscI*/*SalI* for JD983, and *MscI*/*KpnI* for JD908, JD902, and JD900 and run on agarose gels and blotted as described in Example 4. The blots were probed with vector pJY4164. Increasing distance from the *MscI* site to the end of the vector was demonstrated by an increase in the size of the upper band. A faint band in the JD982 lane was observed, likely a result of partial digestion. The 4.7 kb and 4.8 kb bands in JD982 and JD908, respectively, indicate that these mutants contain a duplication of the inserted plasmid.

Insertion of the plasmids results in a duplication of the cloned fragment. Therefore, mutant strains such as JD908, in which the duplicated fragment contains both the 5' end of *cps3S* and the 3' end of *cps3D*, have a full-length copy of *cps3S* downstream of the plasmid insertion. In addition, the full-length copy is contiguous with the 3' end of *cps3D*. Therefore, if *cps3S* had its own promoter, or if one were located in the 3' end of *cps3D*, these insertions should not result in a loss of *cps3S* expression. However, four such insertions have been made in the WU2 chromosome (JD846, JD897, JD898, and JD908), and even with a duplication of up to 450 bp of the 3' end of *cps3D*, a loss of capsule production was observed.

Two more internal insertions in *cps3D* were created. As expected these insertions eliminated capsule production (FIG. 9). However, since *cps3D* and *cps3S* are transcribed as an operon, this result does not prove that *cps3D* is required for capsule synthesis. That fact is demonstrated by the lack of capsule production seen in strains containing non-polar point mutations in *cps3D* (Example 14).

- 114 -

## EXAMPLE 14

In Vitro Polymerization Assay

To evaluate the competence of the mutants to  
5 synthesize type 3 capsule, an in vitro polymerization  
assay was used.

## A. Methods

1. *In vitro polysaccharide synthesis*

10

Type 3 capsular polysaccharide was synthesized and  
quantitated in vitro using a modification of the method  
of Smith, et al., 1961. Crude extracts containing cell  
membranes and cytoplasm were prepared from 200 ml of *S.*  
15 *pneumoniae* cultures harvested at an OD<sub>600</sub> of 0.25 as  
described (Yother and White, 1994), except that cell  
material was concentrated 200-fold, and all steps were  
performed using a thioglycolate buffer (10 mM sodium  
thioglycolate, 5 mM MgSO<sub>4</sub>, 100 mM Tris-HCl pH 8.3) to  
20 stabilize the enzymes (Smith, et al., 1960). The  
digestion of cell wall material by mutanolysin treatment  
was performed in this buffer and 20% sucrose.  
Protoplasts were sonicated three times for 15 s at 35%  
power at 0°C.

25

Polysaccharide synthesis was carried out at 34°C for  
2 h in a 1 ml reaction containing 100 µl of extract, 5 mM  
UDP-glucose, 5 mM UDP-glucuronic acid (where indicated),  
and 1 mM NAD, in the thioglycolate buffer. The reaction  
30 was boiled 1 min then quickly cooled to 25°C in H<sub>2</sub>O.  
Following centrifugation for 30 s at 8160 x g, the type 3  
specific monoclonal antibody 16.3 (Briles, et al., 1981a)  
was added in excess to the supernatant and incubation was  
continued at 37°C for 30 min.

35

The specific antigen-antibody complexes were  
measured at 650 nm in a spectrophotometer, and the amount



- 115 -

of capsule was determined by comparison with a standard curve prepared using purified type 3 polysaccharide purchased from ATCC (Rockville, MD) (Bernheimer, 1953). Reactions were done in triplicate and were standardized to protein content of the crude extract, as determined in duplicate using the Bio-Rad Laboratories (Hercules, CA) protein assay kit.

### B. Results

The spontaneous mutants JD611 and JD619 (*cpsA1* and *cpsA2*), which contain stop mutations in *cps3D*, produce no detectable capsular material. However, both synthesized high molecular weight type 3 polysaccharide in a cell-free system *in vitro* when provided with the nucleotide sugar precursors, i.e., UDP-glucose and UDP-glucuronic acid (Table 7). No capsule was produced by these mutants when UDP-glucuronic acid was omitted from the reaction. These results indicate that these mutants produce no capsule due to the lack of UDP-glucuronic acid and support the conclusion that *Cps3D* is the UDP-glucose dehydrogenase. They also confirm that stop mutations in *cps3D* are not polar on *cps3S*. The increased amount of polysaccharide produced by the WU2 extract (as compared to that produced by that of JD611 or JD619) may be explained by the observation of Smith et al. (1961), that increased amounts of type 3 capsule are produced *in vitro* when a small amount of unpurified polysaccharide is already present in the reaction.

The mutants which contain insertions within *cps3S* (JD902), or between the full-length copies of *cps3D* and *cps3S* (JD908, JD897) were unable to synthesize significant amounts of capsule even with both precursors present. These results emphasize the role of *Cps3S* in capsule synthesis and support the conclusion that *cps3D* and *cps3S* are transcribed as an operon.

- 116 -

The capsule-deficient mutants JD614 and JD692 synthesized only small amounts of additional polysaccharide in the *in vitro* assay. This result is somewhat surprising since JD692, which was shown to contain a missense mutation within the *cps3D* coding region, should still make a functional Cps3S (i.e., the *cps3D* mutation must not be polar since the intact cells are able to synthesize some polysaccharide). The result may suggest that the defective UDP-glucose dehydrogenase in some way interferes with the ability to synthesize the normal polysaccharide. Alternatively, the stability of the *cpsDS* transcript may be altered by the mutation, resulting in a reduced amount of CpsS.

- 117 -

Table 7. *In vitro* capsule synthesis assay.

	strain	Cps phenotype a	UDPGAb	CPS ( $\mu$ g/mg protein)
5				
	JD611	Cps3D <sup>-</sup> S <sup>+</sup>	+	9.8 $\pm$ 0.6
			-	0.9 $\pm$ 0.2
	JD619	Cps3D <sup>-</sup> S <sup>+</sup>	+	5.7 $\pm$ 0.3
			-	0.2 $\pm$ 0.1
10	JD614	Cps3D <sup>+</sup> S <sup>*</sup>	NA <sup>b</sup>	5.4 $\pm$ 0.4 (t <sub>0</sub> ) <sup>c</sup>
			+	5.9 $\pm$ 0.5 (0.5) <sup>c</sup>
	JD692	Cps3D <sup>+</sup> S <sup>*</sup>	NA	4.8 $\pm$ 0.3 (t <sub>0</sub> )
			+	7.0 $\pm$ 1.0 (2.2)
	JD902	Cps3D <sup>+</sup> S <sup>-</sup>	+	1.7 $\pm$ 0.3
15	JD908	Cps3D <sup>+</sup> S <sup>-</sup>	+	1.5 $\pm$ 0.1
	JD897	Cps3D <sup>+</sup> S <sup>-</sup>	+	1.1 $\pm$ 0.1
	WU2	Cps3D <sup>+</sup> S <sup>+</sup>	NA	3.8 $\pm$ 0.2 (t <sub>0</sub> )
			+	16.6 $\pm$ 0.3 (12.8)
			-	16.3 $\pm$ 0.8
20	D39	Cps2 <sup>+</sup>	+	0.5 $\pm$ 0.3

a Capsule phenotypes are based on the *cps3D* and *cps3S* genotypes. - indicates either a stop or insertion mutation (see FIG. 6E, FIG. 6F and FIG. 9 for locations of mutations). \* indicates either a missense or in-frame deletion or insertion in *cps3D* that apparently also affects *cps3S*.

b NA, not applicable.

c For strains which produce capsule *in vivo*, the amount of polysaccharide present at the start of the assay (t<sub>0</sub>) is given, and the amount of polysaccharide produced during the assay is indicated in parentheses.

35

- 118 -

## EXAMPLE 15

Biochemical Pathway

Based on the genetic analysis, the homology of the amino acid sequences of the type-specific genes to the sequences of enzymes of known function, the behavior of the mutants in biochemical and immunochemical assays, and previous biochemical characterizations of type 3 strains (Austrian, et al., 1959; Dillard and Yother, 1994; Smith, et al., 1960; Smith, et al., 1961; Bernheimer, 1953), a pathway for the biosynthesis of type 3 capsular polysaccharide is proposed (FIG. 10). The last of the type-specific genes, *cps3M*, is homologous with phosphoglucosyltransferases from several bacterial species. Even though maintained in the type 3-specific region, *Cps3U* and *Cps3M* may not be required for capsule synthesis, since an insertion internal to *cps3U* (which has a polar effect on *cps3M*) does not result in loss of capsule production (FIG. 7 and FIG. 9), as judged by colony morphology on blood agar medium.

## EXAMPLE 16

The Downstream Non Type-specific  
Flanking Region and Mapping Other Capsule Types

Southern blots of digested chromosomal DNA from strains 2, 3 and 6B and probed with pJD377 was performed as described in Example 4. Faint bands in addition to the band of interest was observed on Southern blots. This was probably due to the detection of fragments containing the *amiA*-like genes which have homology to *plpA*. DNA was either digested with *Bgl*II, *Sac*I or *Hind*III. Other laboratory techniques were as described in Example 1 or Example 8.

Sequence analysis of the 1.2 Kb *Sac*I-*Hind*III fragment (from plasmid JD377) employed in Example 5

- 119 -

contained the 3' end of *cps3M* and the 5' half of a gene with 50% identity to the *S. pneumoniae* *amiA* (SEQ ID NO:6). The *amiA*-like sequence has recently also been identified by Pearce et al. and named *expl* (Pearce, et al., 1993), and subsequently renamed *plpA* (Pearce, et al., 1994). Further Southern hybridizations performed as described in Example 4 showed that the non-type-specific homologous DNA in the 1.2 kb *SacI-HindIII* fragment is *plpA*.

10

A partial copy of a transposase gene was also identified immediately adjacent to and between *cpsM* and *plpA*. Previous findings of repetitive elements linked to the capsule locus suggest that the deletions in this region may be the result of a transposition event, possibly one which introduced the type 3-specific cassette.

15

If, as in type 3, the homologous region is directly adjacent to the type-specific genes in other serotypes, it should be possible to map other type-specific genes using this fragment. This was found to be the case, and the chromosome maps of the capsule regions in strains of types 2, 3, and 6B, from Southern blots, are shown in FIG. 11. It can be seen in FIG. 11 that restriction sites located to the right of the *plpA* fragment are highly conserved in all three strains. The type 3 strain differs slightly in this region due to a deletion of the 5' end of *plpA*. The sites located to the left of *plpA* are divergent among the capsule types. The close linkage of the region to all the necessary type-specific genes for each type, combined with the different restriction maps and the fact that the type 3-specific genes are located directly adjacent to this fragment, suggests that this region contains the type-specific genes in all three capsule types.

20

25

30

35

- 120 -

## EXAMPLE 17

The Upstream Non Type-specific Flanking Region

In order to isolate DNA 5' of the biosynthetic  
5 genes, a 1.8 kb fragment extending from the upstream *SacI*  
sites to just before the *PvuII* site in *cps3D* (nucleotide  
1 through 1802 of FIG. 6D and FIG. 6E, nucleotide 1  
through 934 of SEQ ID NO:4 and nucleotide 1 through 868  
of SEQ ID NO:5) was amplified from the type 3 WU2  
10 chromosome using inverse PCR as described in Example 8.  
All other materials and methods were as described in  
Example 1, 4 and 8.

The 1.8 kb fragment was then used to probe *HindIII*-  
15 digested chromosomal DNA from seven *S. pneumoniae*  
serotypes (2, 3, 5, 6, 8, 9 and 22). The fragment  
hybridized strongly with the expected fragments at 2.2  
and 2.3 kb in the type 3 strain. However, hybridization  
was also observed with fragments of 2.6 and 8 kb, along  
20 with weak hybridization with several other fragments  
(3.0, 3.1, and 4.4 kb). Likewise, each of the strains  
representing other capsule types contained two strongly  
homologous fragments (4.8 and 8.0 kb for types 2, 6B, 8,  
9; 2.2 and 4.8 or 12 for types 5 and 22, respectively)  
25 and at least one weakly homologous fragment (4.4 kb).  
When chromosomal DNAs of types 2, 3, and 6B were digested  
with *PstI*, *PvuII*, or *SacI/HindIII*, and probed with the  
604 bp *SacI-HindIII* fragment (pJD392) upstream of *cps3D*  
(within nucleotides 1 through 610; SEQ ID NO:4, FIG. 6D),  
30 4 to 10 bands were detected in each.

Transformation studies were performed to examine  
linkage of the repeat upstream region to the type-  
specific capsule genes. The plasmid (pJD392) containing  
35 the 604 bp *SacI-HindIII* fragment was introduced into the  
chromosome of the type 3 strain. The insert, located in  
the 2.2 kb *HindIII* fragment (FIG. 7) adjacent to the type

- 121 -

3-specific genes, did not affect capsule production. When the resulting strain was used to transform recipients of types 2 and 6B, greater than 95% of the erythromycin-resistant transformants expressed type 3 capsule. However, when pUD392 was transformed into strains of types 2 and 6B, the plasmid inserted into the 8 kb HindIII fragment, and the type specific genes could not be moved to strains of heterologous types (i.e., 2, 3, or 6B) by transformation and selection for linkage to the erythromycin marker in the insertions. This suggests that this region may be found in more than one location and not necessarily located, adjacent to the cps locus in all other serotypes.

To isolate DNA 5' of the SacI site in the repeat region (SEQ ID NO:4), a 257 nucleotide SacI-MscI fragment internal to the repeat region was used to direct an insertion-duplication event in the type 3 *S. pneumoniae* WU2 chromosome. The flanking 2.3 kb HindIII fragment was then closed from the chromosome, as described in Example 8. The 1.4 kb SacI fragment from within this HindIII fragment was then used to probe HindIII-digested chromosomal DNA from three *S. pneumoniae* serotypes (2, 3, and 6B). The fragment hybridized strongly with the type 2 and 6B strains each contained a strongly homologous band at 0.8, 3.0 and 8.0 kb. The expected 2.3 kb band in the type 3 strain. Less intense bands were also observed at 0.8, 3.0 and 8.0 kb. The other enzymes followed by hybridization with the SacI fragment yielded the results shown in Table 8. The results show that this region is found in only one location in all serotypes examined. The weak bands observed for serotype 3 probably represent hybridization of the 5' end of the repeat region 132 nucleotides upstream of the SacI site (FIG. 7).

- 122 -

TABLE 8. BANDS DETECTED BY SOUTHERN BLOTTING  
Restriction Enzyme used to digest chromosomal DNA.

	<i>Bgl</i> III	<i>Sac</i> I	<i>Sph</i> I
type 2	>12*	4.0, >12	12, >12
5 type 3	>12*	1.4, (weak at 3.5, 8.5, >12)	10 (weak at 12.5, 13)
type 6B	>12*	4.0, 12	12, >12

\* The *Bgl*III fragments were not identical in size.

10 Numbers represent size of bands in kb as observed on  
southern blots carried out as described in Example 4.

DNA in the 1.4 kb *Sac*I fragment was sequenced using  
techniques as described in previous Examples, and can be  
15 seen, alongside the predicted amino acid sequences, in  
FIG. 6A, FIG. 6B and FIG. 6C (SEQ ID NO:1, SEQ ID NO:2,  
SEQ ID NO:3, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 and  
SEQ ID NO:10).

20

## EXAMPLE 18

Capsule Type Expression and Virulence in *S. pneumoniae*

In these studies, isogenic strains expressing the  
type 3 capsule were constructed and the effect on  
25 virulence was determined. Strains of types 2, 5, and 6B  
were used as recipients. The type 2 and 5 strains differ  
in virulence from the type 3 strain in terms of time  
required to cause death (shorter with type 2) and LD<sub>50</sub>  
(lower with type 5). The type 6B strain is of low  
30 virulence in mice. The results showed that expression of  
the type 3 capsule attenuated the virulence of the type 5  
strain, caused the type 6B strain to become highly  
virulent, and had no effect on the type 2 strain. Thus,  
in general, the expression of virulence was correlated  
35 with the type of capsule expressed.



- 123 -

#### A. Methods

##### 1. Transformations, serotyping, ELISAs and restriction enzyme fragment patterns.

Transformations ELISAs and DNA manipulations were performed as described previously in Example 1. All transformants and parental strains were serotyped with capsule type-specific antisera (Statens Seruminstitut, Copenhagen, Denmark) in slide agglutination assays. Genomic DNA, was digested with *HindIII* for 4 h at 37°C and electrophoresed overnight through 0.7% agarose in Tris-borate-EDTA buffer.

##### 2. Analysis of PspA.

Bacteria were grown in CDM containing 2% choline, a condition that causes release of PspA into the culture medium. Filtered, unconcentrated supernatant fluids (20 µl) were electrophoresed in sodium dodecyl sulfate (SDS)-12% polyacrylamide gels. Western blotting (immunoblotting) was performed by using a semidry electroblotter (Bio-Rad Laboratories, Richmond, Calif.), and the blots were processed as described previously (Yother et al., 1992). The PspA-specific monoclonal antibodies XiR278, Xi126, and 2A4 were kindly provided by Larry McDaniel (University of Alabama at Birmingham). Silver staining was performed by using the Silver Stain Kit from Stratagene Cloning Systems, Inc. (La Jolla, Calif.).

##### 3. Characterization of morphology and capsule production.

For average chain length determinations, bacteria were grown in THY to an optical density at 600 nm (OD<sub>600</sub>) of ~0.3. Chain lengths were determined microscopically by using a Petroff-Hauser counting chamber (Aurich C. Thomas Co., Philadelphia, PA). An average of five squares was counted for each strain. Comparisons of average chain lengths were determined by using the two-sample rank test (Zar, 1984).

- 124 -

The number of cells per colony was determined by using bacteria grown on blood agar medium for 18 h at 37°C in 5% CO<sub>2</sub>. A plug containing a single colony was obtained with a sterile Pasteur pipette and then  
5 resuspended in 50 µl of THY. Tenfold serial dilutions were performed in THY and plated on blood agar medium. Plates were incubated overnight at 37°C in 5% CO<sub>2</sub>, and the number of CFU per colony was calculated.

10 Buoyant density determinations were performed by using bacteria grown on blood agar medium or in THY. Bacteria grown on solid medium were harvested by washing each plate with water, centrifuging the suspension, and then resuspending the pellet to an OD<sub>600</sub> of ~0.4 with  
15 water. Ten-milliliter liquid cultures, grown to an OD<sub>600</sub> of ~0.5, were harvested by centrifugation for 10 to 15 min at 8,000 to 16,000 x g. Bacteria were washed twice with water prior to being loaded onto 10-ml, continuous, 0 to 50% Percoll (Pharmacia, Piscataway, N.J.) gradients.  
20 As standards, 5 µl of density marker beads (Pharmacia) ranging in size from 1.033 to 1.076 g/ml, were also loaded. Gradients were centrifuged for 30 min at 8,000 x g with the brake off. A standard curve based on the migration of the marker beads was generated, and the  
25 density of the bacteria was determined by extrapolation.

For determinations of total capsule content, 1.5-ml cultures grown in CDM containing 0.0005% choline were harvested by centrifugation at 8,000 to 16,000 x g for 10  
30 min. Culture supernatant fluids were filtered and saved, and the cells were resuspended in 500 µl of protoplast buffer (20% sucrose, 0.005 M Tris [pH 7.4], 0.0025 M MgSO<sub>4</sub>). Cell sonicates were produced by three cycles of a 10-s pulse, followed by a 10-s incubation on ice, with  
35 a Fisher Sonic Dismembrator model 300 (Fisher Biotechnology, St. Louis, Mo.) with the intensity control

- 125 -

set at 30. Culture supernatant fluids and cell sonicates were stored at ~20°C.

For surface localization assays, 1.5-ml cultures grown to an OD<sub>600</sub> of ~0.5 were heat killed by incubation at 65°C for 30 min. Bacteria were harvested by centrifugation, and culture supernatant fluids were filtered and saved. After the pellets were washed twice with phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na<sub>2</sub>HPO<sub>4</sub> · 7H<sub>2</sub>O, 1.4 mM KH<sub>2</sub>PO<sub>4</sub>), the pellets were resuspended in 1.5 ml of THY. Samples were stored at 4°C.

#### 4. Virulence assays.

The virulence of the type 3 derivatives was compared with that of the parental strains in BALB/ByJ female mice (Jackson Laboratory, Bar Harbor, Maine). Bacteria were grown to the mid-log phase in THY. Samples were diluted serially in sterile lactated Ringer's solution, and 0.2 ml was used to infect mice intraperitoneally (i.p.) or intravenously (i.v.), as indicated. Fifty percent lethal doses (LD<sub>50</sub>s) were determined by the method of Reed and Muench (1938) and compared by Fisher's exact test (Zar 1984). Median times to death were analyzed by using the two-sample rank test (Zar 1984). The P values were determined by using a two-tailed table.

#### B. Results

As described in Table 3, strain JD770 contains a non-destructive insertion in the type 3 capsule locus. The amount and cellular localization of the capsular material produced by JD770 is identical to that of its parent strain WU2. Transformation of JD770 chromosomal DNA, and selection for erythromycin resistance, results in isolates that express the type 3 capsule of the donor but not the capsule of the recipient strain (Example 6). Based on this result, JD770 DNA was used to transform

- 126 -

type 2, 5, and 6B recipients and selected for erythromycin-resistant isolates (see Table 9). All of the type 2 Ery<sup>r</sup> transformants expressed the type 3 capsule but not the type 2 capsule. >95% of the type 5  
5 and type 6B Ery<sup>r</sup> transformants expressed the type 3 capsule but not the capsule of the recipient parent. The remainder of the type 5 and 6B transformants expressed the capsular type of the recipient parent only.

-127-

TABLE 9. BACTERIAL STRAINS

Strain	Derivation <sup>a</sup>	Relevant phenotype	Chain length <sup>b</sup>	Reference
WU2	Clinical isolate	Encapsulated type 3	2.3±0.14	Briles et al. 1981b
JD770	pJD330 <sup>c</sup> x WU2	Encapsulated type 3; Em <sup>r</sup>	2.4±0.19	This study
5 D39	Clinical isolate	Encapsulated type 2	6.2±0.69	Avery et al. 1944
JD803 (2/3) <sup>d</sup>	JD770 x D39	Encapsulated type 3; Em <sup>r</sup>	6.6±0.76	This study
JD804 (2/3)	JD770 x D39	Encapsulated type 3; Em <sup>r</sup>	5.2±0.74	This study
DBL5	Clinical isolate	Encapsulated type 5	9.2±0.80	Yother et al. 1982
TK5010 (5/3)	JD770 x DBL5	Encapsulated type 3; Em <sup>r</sup>	2.1±0.08 <sup>e</sup>	This study
10 TK5011 (5/3)	JD770 x DBL5	Encapsulated type 3; Em <sup>r</sup>	2.4±0.13 <sup>e</sup>	This study
TK5012 (5/3)	JD770 x DBL5	Encapsulated type 3; Em <sup>r</sup>	ND <sup>f</sup>	This study
DBL1	Clinical isolate	Encapsulated type 6B	2.7±1.1	Briles et al. 1992
TK3026 (6B/3)	JD770 x DBL1	Encapsulated type 3; Em <sup>r</sup>	2.4±1.5	This study
TK3028 (6B/3)	JD770 x DBL1	Encapsulated type 3; Em <sup>r</sup>	3.1±3.0	This study

15

<sup>a</sup> Crosses were done by transformation. Each derivative is the result of at least three backcrosses with the parental recipient strain.

<sup>b</sup> The numbers of cells per chain (means ± standard errors) were determined during this study by microscopic examination.

TABLE 9. BACTERIAL STRAINS.....continued

- c pJD330 contains a *Sau3AI* fragment that is part of the type 3 capsule locus. It was constructed in pJY4163 (Yother et al., 1992) which contains an *Em<sup>r</sup>* marker. Insertion of pJD330 into the *S. pneumoniae* chromosome places the *Em<sup>r</sup>* marker in the type 3 locus but does not affect capsule expression.
- d For clarity, strains will be listed by strain number, genetic background, and capsule type expressed. For example, JD803 (2/3) represents a strain with a type 2 genetic background expressing a type 3 capsule.
- e Significantly different from recipient parent strain ( $P < 0.001$ ).
- f ND, not determined.

- 129 -

To produce essentially isogenic strains, two independent transformants from each cross were backcrossed at least three times to their respective parent recipient strains. The final isolates were  
5 examined for restriction enzyme fragment patterns, pneumococcal surface protein A (PspA) expression, capsule expression, and morphological characteristics prior to testing in a mouse virulence model.

10 1. *Restriction enzyme fragment patterns.*

The *Hind*III restriction patterns of the strains used in these studies can be easily distinguished. In all cases, the type 3 derivatives constructed here were found to have the *Hind*III pattern of the recipient strain,  
15 indicating that gross alterations in the genomic DNA content had not occurred and that the parent donor strain JD770 had not been inadvertently re-isolated.

2. *PspA expression.*

20 PspA varies with respect to molecular weight, antigenic determinants, and strain distribution. PspA serotypes and capsular serotypes do not correlate. The strains used in these studies expressed PspAs that had different molecular weights and reacted with different  
25 PspA-specific monoclonal antibodies. In all cases, the PspAs of the type 3 derivatives constructed here were found to have the molecular weight and antibody reactivities of the parent recipient strains.

30 3. *Morphologic characterization and capsule production.*

Microscopic examination revealed that alteration of capsular type had no effect on the chain length of the type 2 and type 6B derivative strains. However, the chain lengths of the type 5 derivatives differed  
35 significantly from that of the type 5 parental strain and

- 130 -

were almost identical to that of the type 3 parent (Table 9).

Morphologically, type 3 strains exhibit large mucoid capsules when grown on blood agar plates, whereas type 2, 5, and 6B strains have small mucoid capsules. The type 3 derivatives of the type 2, 5, and 6B strains had a similar appearance to the type 3 parent on blood agar plates. The increase in colony size compared with that of the recipient parents did not appear to be due to cell number since similar numbers of cells per colony were observed for all of the parent and derivative type 3 strains (data not shown). To examine capsule production, Percoll density gradients and ELISAs were performed. Percoll density gradient centrifugation has been shown previously to differentiate capsular serotypes and amounts by density (Briles et al., 1992). In this assay, all of the derivatives had densities similar to that of the parent type 3 strain and distinct from that of the recipient parent strains (FIG. 12). Thus, all of the derivatives produced cell-associated, surface-localized type 3 capsule in amounts similar to that of the type 3 parent. The total amounts of capsule material produced, i.e., both cell associated and released, were determined in ELISAs to be similar for both the type 3 parent and each of the derivatives (FIG. 18). ELISAs were also used to confirm that the amounts of surface-accessible capsule were similar in the type 3 parent and the derivatives.

#### 4. Virulence of type 3 derivatives.

To assess the effect of alteration of capsule type on virulence, BALB/ByJ female mice were infected i.p. or i.v. with the type 3 derivatives and parent strains. Strain JD770, which contains the nondestructive erythromycin resistance marker in the type 3 capsule locus, did not differ from its parent type 3 strain WU2



- 131 -

in terms of median time to death (52 versus 49.5 h, i.p.) or LD<sub>50</sub> (75 versus 50 CFU, i.p.;  $1 \times 10^5$  versus  $2 \times 10^5$  CFU, i.v.). Thus JD770 was used in subsequent studies for comparisons with the type 3 derivatives. As  
5 expected, the recipient parent strains were significantly different from JD770 with respect to time to death or LD<sub>50</sub>s (FIG. 14). Expression of the type 3 capsule had no apparent effect on the virulence of the type 2 recipient strain; i.e., the time required to cause death was not  
10 significantly different from that of the type 2 parent but was significantly different from that of the type 3 parent (FIG. 14A). However, alteration of capsular type had dramatic effects on the virulence of the type 5 and 6B strains. In contrast to the highly virulent type 5  
15 parental strain (LD<sub>50</sub>, ~10 CFU) and the virulent type 3 parental strain (LD<sub>50</sub>, ~ $10^5$ ), the type 3 derivatives were not virulent even at doses of  $10^6$  CFU (FIG. 14B). Switching of the type 6B capsule to type 3 resulted in a  
20 reduction of the LD<sub>50</sub> from  $>1 \times 10^6$  to  $\sim 6 \times 10^3$  CFU, a value that was similar to but still greater than the  $7.5 \times 10^2$  value observed for the type 3 parent strain (FIG. 14C).

These results may be indicative of the role other  
25 factors play in pneumococcal virulence. For example, the type 5 capsule may represent one that results in high virulence with few other factors required, whereas the type 3 capsule may require the presence of other factors to be highly virulent. The introduction of the type 3  
30 capsule into the type 5 genetic background may thus result in the expression of a virulent capsule but, in the absence of other necessary factors, in an avirulent strain. The increase in virulence of the type 6B strain suggests that the type 3 capsule is probably more  
35 virulent than the type 6B. However, its failure to become as virulent as the type 3 parent is suggestive of

- 132 -

a lack of other virulence factors in the 6B background. The type 2 recipient was only slightly more virulent than the type 3 donor and no significant change was noted in the virulence of its derivatives (FIG. 12). This result  
5 may suggest that the type 2 and type 3 capsules are of equal virulence and that the "accessory factors" necessary for full virulence are present in both strains.

Whether the decrease in virulence of type 5  
10 derivatives is related to the alteration of cell chain length is not known. Clearly, the parent type 3 strain is highly virulent with a similar chain length. The alteration in chain length may reflect a general change in the surface structure of the type 5 strains possibly  
15 resulting from the change in capsule expression. Because the strains constructed were transformed with chromosomal DNA the inventors cannot rule out the possibility that determinants closely linked to the capsule locus are affecting the outcome of these studies. However, because  
20 several backcrosses were performed, and because independent isolates exhibited identical characteristics, it is unlikely that unlinked determinants are responsible for the results.

25

#### EXAMPLE 19

##### Increased Virulence of *S. pneumoniae* type 6B by Inactivation of *plpA*.

In Example 18, the introduction of the type 3-  
30 specific cassette and linked genes into an avirulent type 6B strain resulted in expression of the type 3 capsule and an increase in virulence. To more clearly define the contribution of the capsular serotype to the virulence of *S. pneumoniae*, insertion-duplication mutagenesis was used  
35 to insert an erythromycin marker adjacent to the type 6B-specific capsule cassette in the 3' flanking region.

- 133 -

Surprisingly, introduction of this insertion resulted in an increase in virulence of the type 6B strain (derivative LD<sub>50</sub> of 10<sup>3</sup> versus parental LD<sub>50</sub> of >10<sup>6</sup>, intraperitoneal). This enhancement of virulence could be attributed to the 1.2 kb *SacI-HindIII* fragment from the type 3 strain WU2 (pJD377) that was used to direct the erythromycin marker into the type 6B chromosome. Transformation of the wild type 6B strain with the *SacI-HindIII* fragment alone, followed by intraperitoneal infection of the transformation mixture into mice resulted in death in less than 24 hours. Identical results were obtained using multiple smaller fragments of the *SacI-HindIII* fragment. JD377 (Example 5) comprises the 3' end of *cps3M* and part of the gene *plpA*. The fragments used contained mutations that, like the original insertion, resulted in a defective *plpA* in the type 6B strain. These data suggest that avirulence of type 6B observed via the intraperitoneal route is due to expression of *plpA*, and that the increase in virulence of the type 6B strain expressing the type 3 capsule is the result of inactivation of the linked *plpA* and not expression of the type 3 capsule.

\* \* \*

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the composition, methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically

- 134 -

and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are  
5 deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

#### REFERENCES

10 The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

15 ACIP, "Pneumococcal polysaccharide vaccine," *Morbid. Mortal. Weekly*, 30:410-419, 1981.

ACIP, "Pneumococcal polysaccharide vaccine," *Morbid. Mortal. Weekly*, 38:64-76, 1989.

20 Albright et al., "A 13-amino acid peptide in three yeast glycosyltransferases may be involved in dolichol recognition." *Proc. Natl. Acad. Sci. USA*, 86:7366, 1989.

25 Appelbaum, "Antimicrobial resistance in *Streptococcus pneumoniae*: an overview," *Clin. Infect. Dis.*, 15:77-83, 1992.

30 Atkinson and Long, "Homology of *Rhizobium meliloti* NodC to polysaccharide polymerizing enzymes." *Mol. Plant-Microbe Interact.* 5:439, 1992.

Austrian, "Pneumococcal infections," In: Germanier R. ed.  
35 Bacterial vaccines, New York: Academic Press, Inc., 257-288, 1984.

- 135 -

- Austrian et al., "Simultaneous production of two capsular polysaccharides by pneumococcus II. The genetic and biochemical bases of binary capsulation," *J. Exp. Med.*, 110:585-602, 1959.
- 5 Ausubel et al., *Current Protocols in Molecular Biology*, New York: Wiley, 1987.
- 10 Avery et al., Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III," *J. Exp. Med.*, 79:137-158, 1944.
- 15 Bastin et al., "Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length," *Mol. Microbiol.*, 7:725-734, 1993.
- 20 Bernheimer, A.W., "Synthesis of type III pneumococcal polysaccharide by suspensions of resting cells." *J. Exp. Med.*, 97:591, 1953.
- 25 Bernheimer et al., "Mutation in *Pneumococcus* type III affecting multiple cistrons concerned with the synthesis of capsular polysaccharide." *J. Bacteriol.*, 96:1099, 1968.
- 30 Bernheimer and Wermundsen, "Unstable binary capsulated transformants in pneumococcus," *J. Bacteriol.*, 98:1073-1079, 1969.
- 35 Bernheimer et al., "Qualitative differences in the behavior of pneumococcal deoxyribonucleic acids

- 136 -

transforming to the same capsular type," *J. Bacteriol.*, 93:320-333, 1967.

- 5       Bernheimer and Wermundsen, "Homology in capsular transformation reactions in *Pneumococcus*," *Molec. Gen. Genet.*, 116:68-83, 1972.
- 10       Birnboim and Doly, "A rapid alkaline extraction procedure for screening recombinant plasmid DNA," *Nucleic Acids Res.*, 7:1513-1523, 1979.
- 15       Branconier and Odeberg, "Granulocyte phagocytosis and killing of virulent and avirulent serotypes of *S. pneumoniae*," *J. Lab. Clin. Med.* 100:276-287 1982.
- 20       Briese and Hakenbeck, "The target of penicillin. The murein sacculus of bacterial cell walls, architecture, and growth," Berlin, Walter de Gruyter, 1983.
- 25       Briles et al., "A strong association between capsular type and mouse virulence among human isolates of *Streptococcus pneumoniae*," *Infect. Immun.*, 60:111-116, 1992.
- 30       Briles et al., "Genetic control of susceptibility to pneumococcal infection," *Curr. Top. Microbiol. Immunol.*, 124:103-120, 1986.
- 35       Briles et al., "Mouse IgG3 antibodies are highly protective against infection with *Streptococcus pneumoniae*," *Nature*, 294:88-90, 1981a.
- Briles et al., "Antiphosphocholine antibodies found in normal mouse serum are protective against

- 137 -

intravenous infection with type 3 *Streptococcus pneumoniae*," *J. Exp. Med.*, 153:694-705, 1981b.

5        Brosius and Holy, "Regulation of ribosomal RNA promoters with a synthetic lac operator." *Proc. Natl. Acad. Sci. USA* 81:6929, 1984.

10        Burke et al., "Pneumococcal bacteremia: review of 111 cases, 1957-1959, with special references to cases of undetermined focus," *Am. J. Dis. Child.*, 121:353-359, 1971.

15        Campbell, in *Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology* Vol. 13, Burden and Von Knippenberg, Eds. pp. 75-83, Amsterdam, Elsevier, 1984

20        Centers for Disease Control, "Bacterial meningitis and meningococemia - United States, 1978," *Morbid. Mortal. Weekly Rep.*, 28:277-279, 1978.

25        Coffey et al., "Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*," *Molec. Microbiol.*, 5:2255-2260, 1991.

30        Crain et al., "Pneumococcal surface protein A (PspA) is serologically variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*," *Infect. Immun.*, 58:3293-3299, 1990.

35        Dawson, M.H., "The transformation of pneumococcal types. II. The interconvertibility of type-specific *S* pneumococci." *J. Exp. Med.*, 51:123, 1930.

- 138 -

- Dawson and Sia, "In vitro transformation of pneumococcal types. I. A technique for inducing transformation of pneumococcal types in vitro," *J. Exp. Med.*, 54:681-699, 1931.
- 5 DeAngelis et al., "Isolation of a *Streptococcus pyogenes* gene locus that directs hyaluronan biosynthesis in acapsular mutants and in heterologous bacteria." *J. Biol. Chem.*, 268:14568, 1993a.
- 10 DeAngelis et al., "Molecular cloning, identification, and sequence of the hyaluronan synthase gene from group A *Streptococcus pyogenes*." *J. Biol. Chem.*, 268:19181, 1993b.
- 15 Debelle et al., "The *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium* NodC proteins are homologous to yeast chitin synthases." *Mol. Plant-Microbe Interact.*, 5:443, 1992.
- 20 Deretic, et al., "*Pseudomonas aeruginosa* infection in cystic fibrosis: nucleotide sequence of the *algD* gene." *Nucleic Acids Res.*, 15:4567, 1987.
- 25 Dillard and Yother, "Analysis of *Streptococcus pneumoniae* sequences cloned into *Escherichia coli*: effect of promoter strength and transcription terminators," *J. Bacteriol.*, 173:5105-5109, 1991.
- 30 Dillard and Yother, "Genetic and molecular characterization of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 3." *Molec. Microbiol.*, 12:959, 1994.
- 35 Dougherty and van de Rijn, "Molecular characterization of *hasA* from an operon required for hyaluronic acid



- 139 -

synthesis in group A streptococci." *J. Biol. Chem.*,  
269:169, 1994.

5 Dougherty and van de Rijn, "Molecular characterization of  
hasB from an operon required for hyaluronic acid  
synthesis in Group A streptococci," *J. Biol. Chem.*,  
268:7118-7124, 1993.

10 Effrussi-Taylor, "Genetic aspects of transformation of  
pneumococci," *Cold Spr. Harb. Symp. Quant. Biol.*,  
16:445-456, 1951.

15 Fine, "Pneumococcal type-associated variability in  
alternate complement pathway activation," *Infect.*  
*Immun.*, 12:772-778, 1975.

20 Finland and Barnes, "Changes in occurrence of capsular  
serotypes of *Streptococcus pneumoniae* at Boston City  
Hospital during selected years between 1935 and  
1974," *J. Clin. Microbiol.*, 5:154-166, 1977.

25 Frosch et al., "Molecular characterization and expression  
in *Escherichia coli* of the gene complex encoding the  
polysaccharide capsule of *Neisseria meningitidis*  
group B," *Proc. Natl. Acad. Sci. USA*, 86:1699-1673,  
1989.

30 Garcia et al., "Cloning of a gene involved in the  
synthesis of the capsular polysaccharide of  
*Streptococcus pneumoniae* type 3," *Mol. Gen. Genet.*,  
239:188-195, 1993.

Gefter et al., *Somatic Cell Genet.* 3:231-236 (1977)

35 Genetics Computer Group. Program Manual for the GCG  
Package, Version 7, 1991.

- 140 -

Giebink et al., "Opsonic requirements for phagocytosis of *S. pneumoniae* types VI, XVII, XXIII, and XXV." *Infect. Immun.* 18:291-297, 1977.

- 5     Goding, in *Monoclonal Antibodies: Principles and Practice*, 2d ed., Orlando, Fla., Academic Press, 1986, pp. 60-61, 65-66, 71-74, 1986.

- 10     Gordon et al., "Ligand-receptor interactions in the phagocytosis of virulent *Streptococcus pneumoniae* by polymorphonuclear leukocytes," *J. Infect. Dis.*, 154:619-626, 1986.

- 15     Gray and Dillon, "Clinical and epidemiologic studies of pneumococcal infection in children," *Pediatr. Infect. Dis.*, 5:201-207, 1986.

- 20     Harr et al., "Search algorithm for pattern match analysis of nucleic acid sequences." *Nuc. Acids Res.*, 11:2943, 1983.

- 25     Hostetter, "Serotypic variation among virulent pneumococci in deposition and degradation of covalently bound C3b: implications for phagocytosis and antibody production," *J. Infect. Dis.*, 153:682-693, 1986.

- 30     Jacobs et al., "Emergency of multiply resistant pneumococci," *N. Engl. J. Med.*, 299:735-740, 1978.

- Janoff et al., "Pneumococcal disease during HIV infection: epidemiologic, clinical, and immunologic perspectives," *Ann. Intern. Med.*, 117:314-324, 1992.

- 141 -

- John et al., "Transmembrane orientation and receptor-like structure of the *Rhizobium meliloti* common nodulation protein NodC." *EMBO J.*, 7:583, 1988.
- 5 Johnson et al., "Immunogold localization of the NodC and NodA proteins of *Rhizobium meliloti*." *J. Bacteriol.*, 171:4583, 1989.
- 10 Johnston and Sell, "Septicemia in infants and children," *Pediatrics.*, 34:473-479, 1964.
- Kado and Liu, "Rapid procedure for detection and isolation of large and small plasmids," *J. Bacteriol.*, 145:1365-1373, 1981.
- 15 Katsube et al., "Expression in *Escherichia coli* of UDP-glucose pyrophosphorylase cDNA from potato tuber and functional assessment of the five lysyl residues located at the substrate binding site."
- 20 *Biochemistry*, 30:8546, 1991.
- Kazuta et al., "Identification of lysyl residues located at the substrate-binding site in UDP-glucose pyrophosphorylase from potato tuber: affinity labeling with uridine di- and triphosphopyridoxals,"
- 25 *Biochemistry*, 30:8541, 1991.
- Kelleher et al., "Oligosaccharyl transferase activity is associated with a protein complex composed of
- 30 ribophorins I and II and a 48 kd protein." *Cell*, 69:55, 1992.
- Knecht et al., "Some biologic properties of pneumococcus type 37 and the chemistry of its capsular polysaccharide," *J. Exp. Med.*, 132:475-485, 1970.
- 35

- 142 -

- Koch and Dennison, "Office visits to pediatricians,"  
National Ambulatory Medical Care Service, National  
Center for Statistics, 1974.
- 5 Kohler and Milstein, *Nature* 256:495-497, 1975.
- Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976.
- 10 Kroll et al., "Common organization of chromosomal loci  
for production of different capsular polysaccharides  
in *Haemophilus influenzae*," *J. Bacteriol.*,  
171:3343-3347, 1989.
- 15 Kyte and Doolittle, "A simple method for displaying the  
hydropathic character of a protein." *J. Mol. Biol.*,  
157(1):105-132, 1982.
- 20 Landesman et al., "Infections in children with sickle  
cell anemia. Special reference to pneumococcal and  
salmonella infections," *Am. J. Pediatr. Hematol  
Oncol.*, 4:407-18, 1982.
- 25 Langvad-Nielson, "Change of capsule in the pneumococcus,"  
*Acta Path et Microbiol. Scand.*, 21:362-369, 1944.
- Lerouge et al., "Symbiotic host specificity of *Rhizobium  
meliloti* is determined by a sulphated and acylated  
glucosamine oligosaccharide signal." *Nature*,  
344:781, 1990.
- 30 MacLeod, "Pneumococci," In: *Bacterial and mycotic  
infections of man*, 391-411, Dubose and Hirsch eds.,  
Lippincott, Philadelphia, PA, 1965.
- 35 Morrison et al., "Isolation of transformation-deficient  
*Streptococcus pneumoniae* mutants defective in

- 143 -

control of competence, using insertion-duplication mutagenesis with the erythromycin resistance determinant of pAMB1," *J. Bacteriol.*, 159:870-876, 1984.

5

Muckerman et al., "Transformation of restriction endonuclease phenotype in *Streptococcus pneumoniae*," *J. Bacteriol.*, 152:183-190, 1982.

10 Mulligan and McClure, "Analysis of the occurrence of promoter-sites in DNA." *Nuc. Acids Res.*, 14:109, 1986.

15 Overturf et al., "Bacterial meningitis and septicemia in sickle cell disease," *Am. J. Dis. Child.*, 131:784-7, 1977.

Pearce et al., "Genetic identification of exported proteins in *Streptococcus pneumoniae*." *Molec. Microbiol.*, 9:1037, 1993.

20

Pearce et al., "Peptide permeases modulate transformation in *Streptococcus pneumoniae*." *Molec. Microbiol.*, 12:881, 1994.

25

Powars, "Natural history of sickle cell disease--the first ten years," *Semin. Hematol.*, 12:267-85, 1975.

30 Powars et al., "Pneumococcal septicemia in children with sickle cell anemia. Changing trend of survival," *JAMA*, 245:1839-42, 1981.

35 Radloff et al., "A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells," *Proc. Natl. Acad. Sci. USA*, 57:1514-1521, 1967.

- 144 -

- Ravin, "Reciprocal capsular transformation of pneumococci," *J. Bacteriol.*, 77:296-309, 1959.
- 5       Ravin, "Linked mutations borne by deoxyribonucleic acid controlling the synthesis of capsular polysaccharide in pneumococcus," *Genetics* 45:1387-1403, 1960.
- 10       Reed and Muench, "A simple method of estimating fifty percent endpoints," *Am. J. Hyg.* 27:493-497, 1938.
- 15       Reeves and Goebel, "Chemoimmunological studies on the soluble specific substance of pneumococcus. V. The structure of the type III polysaccharide," *J. Biol. Chem.*, 139:511-519, 1941.
- 20       Ridley, et al., "Mechanism of action of uridine diphosphoglucose dehydrogenase: evidence for a second reversible dehydrogenation step involving an essential thiol group." *J. Biol. Chem.* 250:8761, 1975.
- 25       Roberts et al., "Common organization of gene clusters for production of different capsular polysaccharides (K antigens) in *Escherichia coli*," *J. Bacteriol.*, 170:1305-1310, 1988.
- 30       Rubens, et al., "Identification of cpsD, a gene essential for type III capsule expression in group B streptococci." *Molec. Microbiol.*, 8:843, 1993.
- Sambrook et al., "Molecular cloning: A laboratory manual," *Cold Springs Harbor Laboratory*, Cold Springs Harbor, NY, 1989.

- 145 -

- Saunders and Guild, "Properties and transforming activities of two plasmids in *Streptococcus pneumoniae*," *Mol. Gen. Genet.*, 128:283-290, 1980.
- 5 Schiller et al., "UDP-glucose dehydrogenase from *Escherichia coli*. Purification and subunit structure," *Biochim. Biophys. Acta*, 453:418-425, 1976.
- 10 Schutzbach, et al., "The purification and characterization of recombinant yeast dolichol-phosphate-mannose synthase. Site-directed mutagenesis of the putative dolichol recognition sequence." *J. Biol. Chem.*, 268:24190, 1993.
- 15 Shoemaker and Guild, "Destruction of low efficiency markers is a slow process occurring at a heteroduplex stage of transformation," *Molec. Gen. Genet.*, 128:283-290, 1974.
- 20 Smith, et al., "Uridine pyrophosphoglucose dehydrogenase in capsulated and non-capsulated strains of pneumococcus type I." *J. Gen. Microbiol.*, 22:265, 1960.
- 25 Smith, et al., "The formation of uridine pyrophosphoglucuronic acid from uridine pyrophosphoglucose by extracts of a noncapsulated strain of pneumococcus." *Biochim. Biophys. Acta.*, 28:211, 1958.
- 30 Smith et al., "The synthesis of type III pneumococcal capsular polysaccharide from uridine nucleotides by a cell-free extract of *Diplococcus pneumoniae* type III," *J. Biol. Chem.*, 235:1876-1880, 1960.
- 35

- 146 -

- Smith, et al., "Biosynthesis of pneumococcal capsular polysaccharides: I. Properties of the system synthesizing type III capsular polysaccharide," *J. Biol. Chem.*, 236:2179, 1961.
- 5 Sorensen, et al., "Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunochemical methods," *Microbial. Pathog.*, 8:325, 1990.
- 10 Stephens et al., "Classical and alternative complement pathway activation by pneumococci," *Infect. Immun.*, 17:296, 1977.
- 15 Taylor, "Additive effects of certain transforming agents from some variants of pneumococcus," *J. Exp. Med.*, 89:399-424, 1949.
- 20 Tilghman et al., "Cloning specific segments of the mammalian genome: bacteriophage lambda containing mouse globin and surrounding gene sequences," *Proc. Natl. Acad. Sci. USA*, 74:4406-4410, 1977.
- 25 van Dam et al., "Immunogenicity and immunochemistry of *Streptococcus pneumoniae* capsular polysaccharides," *Antonie van Leeuwenhoek*, 58:1-47, 1990.
- 30 Versalovic et al., "Penicillin-resistant *Streptococcus pneumoniae* strains recovered in Houston: identification and molecular characterization of multiple clones," *J. Infect. Dis.*, 167:850-856, 1993.
- 35 Walter et al., "Extension of the separation of types among pneumococci: description of 17 types in



- 147 -

addition to types 1 to 32 (Cooper)," *J. Immunol.*,  
41:279, 1941.

5 Wierenga, et al., "Prediction of the occurrence of the  
ADP-binding bab-fold in proteins, using an amino  
acid sequence fingerprint," *J. Mol. Biol.*, 187:101,  
1986.

10 Winkelstein et al., "Activation of C3 via the alternative  
complement pathway results in fixation of C3b to the  
pneumococcal cell wall," *J. Immunol.*, 124:2502-2506,  
1980.

15 Winkelstein et al., "The role of the capsular  
polysaccharide in the activation of the alternative  
pathway by the pneumococcus," *J. Immunol.*,  
116:367-370, 1976.

20 Wood and Smith, "The inhibition of surface phagocytosis  
by the capsular 'slime layer' of pneumococcus type  
III," *J. Exp. Med.*, 90:85-96, 1949.

25 Yother and White, "Novel surface attachment mechanism of  
the *Streptococcus pneumoniae* protein PspA," *J.*  
*Bacteriol.*, 176:2976, 1994.

30 Yother et al., "Protection of mice from infections with  
*Streptococcus pneumoniae* by anti-phosphocholine  
antibody," *Infect. Immun.*, 36:184-188, 1982.

35 Yother et al., "Truncated forms of PspA that are secreted  
from *Streptococcus pneumoniae* and their use in  
functional studies and cloning of the *pspA* gene," *J.*  
*Bacteriol.*, 174:610-618, 1992.

- 148 -

Yother et al., "Transformation of encapsulated  
*Streptococcus pneumoniae*," *J. Bacteriol.*,  
168:1463-1465, 1986.

- 5 Zar, "Biostatistical analysis," p. 718. Prentice Hall,  
Inc., Englewood Cliffs, N.J. 1984.

Zwahlen et al., "The molecular basis of pathogenicity in  
*Haemophilus influenzae*: comparative virulence of  
10 genetically-related capsular transformants and  
correlation with changes at the capsulation locus  
*cap*," *Microbial. Pathog.*, 7:225-235, 1989.

- 149 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5

## (i) APPLICANT:

(A) NAME: UAB RESEARCH FOUNDATION

(B) STREET: 701 SOUTH 20TH STREET  
SUITE 1120G

10

ADMINISTRATION BUILDING

(C) CITY: BIRMINGHAM

(D) STATE: ALABAMA

(E) COUNTRY: UNITED STATES OF AMERICA

(F) POSTAL (ZIP) CODE: 35294-0111

15

(ii) INVENTORS: YOTHER, Janet  
DILLARD, Joseph P.(iii) TITLE OF INVENTION: *STREPTOCOCCUS PNEUMONIAE*  
CAPSULAR POLYSACCHARIDE  
GENES AND FLANKING REGIONS

20

(iv) NUMBER OF SEQUENCES: 20

25

## (v) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Arnold, White &amp; Durkee

(B) STREET: P.O. Box 4433

(C) CITY: Houston

(D) STATE: TX

30

(E) COUNTRY: United States of America

(F) ZIP: 77210

## (vi) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

35

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

- 150 -

(D) SOFTWARE: PatentIn Release #1.0, Version  
#1.30

(vii) CURRENT APPLICATION DATA:

- 5 (A) APPLICATION NUMBER: UNKNOWN  
(B) FILING DATE: CONCURRENTLY HEREWITH  
(C) CLASSIFICATION: UNKNOWN

(viii) PRIOR APPLICATION DATA:

- 10 (A) APPLICATION NUMBER: US 08/243,546  
(B) FILING DATE: 16-MAY-1994  
(C) CLASSIFICATION: UNKNOWN

(ix) ATTORNEY/AGENT INFORMATION:

- 15 (A) NAME: Parker, David L.  
(B) REGISTRATION NUMBER: 32,165  
(C) REFERENCE/DOCKET NUMBER: AMCY018P--

(x) TELECOMMUNICATION INFORMATION:

- 20 (A) TELEPHONE: (512) 418-3000  
(B) TELEFAX: (713) 789-2679  
(C) TELEX: 79-0924

25 (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 base pairs  
(B) TYPE: nucleic acid  
30 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "DNA"

-151-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAGCTCGGTA TTTTITGGAA CGTGATTAG TTCATGTAGT TGCAAGTGAC ATGCACAATT 60

5 TAGACAGTAG ACCTCCATAT ATGCAACAGG CATATGATAT CATTGCTAAG AAATATAGAG 120

CGAAAAAAGC GAAAGAACTT TTTGTAGATA ATCCAGAAA AATTATAATG GATCATTAAT 180

TAGGAGAAAA TATGAAGGAA CAAACACCTT TGGAAATCGA TGTATTGCAG TATTCAGAGC 240

10 TTATTGGAAG AAGTGTCAAT TTATTAGTGG CATTATACCTT CTTCAGTTGC TTTTTCCTAC 300

(2) INFORMATION FOR SEQ ID NO:2:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 261 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "DNA"

-152-

(\*i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

5	CTTCCGTATG CCGCGTTCCT GCAAGACCCC ATCGGCTGGC TCTTTGATAG AGTAGCTGCC	60
	CAAAAAATTA TCAGTATTAC TCGTGCTGAT GTGGCACACT GGAGGAGCAA GACCGCCGAT	120
	ATCACCGCTT CGCCAAATAA TAAACGCAAT AACTAATTG GTTTTTTGGC ATTTTTTTATT	180
10	GGAAGTAGTG TCATAGTTCT TCTTCTTGAA CTTTTGGACA CTCATGTGAA ACGTCCGGAA	240
	GATATCGAAG ATACACTGCA G	261

15 (2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 262 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

. . . . .

-153-

(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

5	CAGCACTTCT CATCTGGCAC AGCTGATTTA TCTCACGGCT TATGTGATAC AAATATTGAA	60
	AATTTATTG TAGTTCAATC GGGATCTGTA TCACCAAACC CTACAGCCTT GTCACAAAGC	120
	AAAAATTTTG TGGTTATGGT AAAGCTTTT TCAAAAAGAGG TCAGTATATT GAGTTGGTGG	180
10	AAACGATAAA ATACATAATA TTTTATTCCT TTGGTTATCA AATTAGCCCC TCCTGAAGCT	240
	CCCCAATTGA CGGCTTGAGC TC	262

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 934 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

20

-154-

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

5  
GAGCTCCAAT CAAAGGGTGT TTGTACACTT TTGACAGAG GGTAAATCGCT AGAGGACAGC 60  
AAACAGCCAT AGTAGTGAAA AATCCAGACA CCTAAAGCAG ACAAAGGGT TGCCATCAGG 120  
10 TATAAAATCA TGTAGAGGGC GTTAGGGTAG GTGCGTGTGC GGTAGAGAAT GTGTTGAGCC 180  
AAACATCAA GAGTACCGTT AGTTATTGCA AAGTTATAAA AGAGAGAGAC GCTAAAAATG 240  
GTAAAAAGAG TGAGTTGGCC AAAATGAAGA AGTTCITTGG GGCITTAATCC CATGAGAGTG 300  
15 GTTGCGATGA GGTAAAGAAA AGCAATAGCC AGCAGGTCAA TATTGATTTT GGTGCGGTAA 360  
CCAATTCCAA TGGCTAGAGC AATGGCGCTA ATCATTATTA AATGAATCAT TGAATTGTCC 420  
20 TTTAGTTAGA ATATAGAAAG AGGATAGATT GAAGTTCGAG AATACTGGGT GTCTTCTGAT 480  
GTTAAGTGGT TGTGTCAAAA CCATCCCCAA TGGCATAAAT TGCAGTGGAA TTGGAATGAC 540



-155-

TCGTAATACT ATTGATAGAG ATGGTATTAC AAGTCAAGAC GTTCGCTATT TTATCTTTAA 600  
CTTTAAGCTT AGTGTAGTAG CCTTTTGCCA TAGTGTTTGA GGTCATTGGT CAGCAGAAAG 660  
TATGCATTGG TTATTGGATG TGGTTTATCG TGAAGACCAT CATCAGACCC TGGATAAACG 720  
AGCTGCTTTT AACCTTAATC TTATCGAAAA ATGTGTTTAT ATTTTCTAAA AGTGATGGTA 780  
TTTTCTAAAA AAGACCTCAG CTATCGATGC AAACAACGGT ATATCTCTGT ACATTGGAA 840  
GATTATTGG AACACAGAGGT TAGGAAAGTA ATCAGTTAAC GGGATATCTT TTCAAAGCTG 900  
ATACTAAGGC ACAAAAAAAA GTTTGATATT CCCC 934

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4951 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

15

10

-156-

(ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
5          60
TTGACAATAG ATAAAATTAT TATATAATTA AACTATTGCT TTTTAAATAA AGTGAGAATA

          120
TTAATAATGC AGAGAAAGAG GACTGTAGTA AAATGAAAAT TGCCATTGCA GGAAGTGGTT

          180
ATGTAGGTCT GTCTTTAGCG GTGCTACTAG CTCAGCATCA TGAAGTTAAG GTCATTGATG

          240
TTATAAAAGG TAAGGTAGAG TCGATAAACA ATAGAAAATC TCCAATTAAG GATGAGGCGA

          300
TTGAGAAATA CTTAGTTGAA AAAGAGTTGA ATCTTGAAGC CTCCTTAGAT CCTGCACACG

          360
TTTATAAAGA CGTGGAGTAT GCTATTATTG CTACTCCGAC TAATTATGAT GTAGACTTAA

          420
ATCAGTTTGA TACATCTTCA GTTGAAGCTG CTATCAAGAC TTGTATGGAA TATAATGATA

          480
CTTGTTACAAT CGTAATCAAA AGTACTATTG CTGAAGGGTA TACTAAAGAA GTGAGGGGAAA

          540
AGTTTAATAC AGATCGTATT ATTTTTTCTC CAGAGTTTCT ACGTGAATCC AAAGCTTTAT
```

-157-

ATGATAAATT GTATCCATCT AGAATTGTTG TAGGAAC TGA TTTGGATGAT TCTGAGTTAA 600  
CAAAAAGAGC ATGGCAGTTT GCAGATCTAC TTAAAGGTGG AGCTATTAAAG GAAGAGGTTTC 660  
5 CGATACTGGT TGTGCTTTT AATGAAGCAG AGGTGCAAA ATTGTTTAGT AACACTTACT 720  
TGGCAACTCG CGTAGCTTAT TTAAATGAGA TAGATACATA TAGCGAGGTA AAAGGGCTTA 780  
ATCCCAAGAC AATTATTGAT ATTGTTTGT ATGATCCTAG AATTGGATCA GACTATAATA 840  
ACCCTAGCTT TGGTTACGGA GGGTATTGCT TACCAAAAGA CACAAAGCAA TTGAAAGCAA 900  
GTTT TAGGGA TGTTCCTGAA AATCTGATTA CAGCTGTGGT GCAATCTAAT AAAACAAGAA 960  
15 AAGATTATAT AGCTGGAGCT ATTCTAGCTA AACAACTAG TGTGTAGGT ATTTATAGAT 1020  
TAATTATGAA ATCTGATTCT GATAATTTTC GTTCTAGTGC TGTTAAGGGA GTTATGGAAC 1080  
GTTTGGACAA TTATGGTAAA GAAATTGTTA TTTACGAACC TACTATTGAG TGTGATACTT 1140  
20 TTATGGGATA CAGAGTAAT AAATCTTTAG ATGAATTAA GAATATTCT GACATTGTTG 1200  
TAGCGAATCG TATGAACGAT GATTTAAGG ATATACAAGA AAACTCTAT ACACGGGATT 1260

-158-

1320 TATTTGGCAG AGAATAACGG GAAATAAATT TTATGTATAC ATTTATTTTA ATGTTGTTTG  
1380 ATTTTTTCA GAATCATGAT TTTCATTCTT TTATGTTGTT TTTTGTCTTT ATTCTTATTC  
1440 5 GTTGGGCGGT TATATATTTT CATGCTGTCA GATATAAGTC CTACAGTTGT AGTGTAAGTG  
1500 ATGAGAAAGTT ATTTAGTTCT GTAATTATCC CTGTCGTGGA TGAACCACTT AATCTTTTTG  
1560 AAAGTGACT GAATAGAATT TCCAGACATA AACCATCCGA AATTATTGTG GTTATTAAACG  
1620 GCCCAAAAA CGAGAGACTT GTAAAACTTT GTCATGATTT TAATGAAAAA TTAGAAAAATA  
1680 ATATGACTCC AATTCAATGT TATTACACTC CTGTTCCCTGG CAAGAGAAAT GCTATCCGCG  
1740 15 TTGGGCTGGA GCATGTGGAT TCGCAGAGTG ATATTACAGT TCTAGTAGAT AGTGATACAG  
1800 TATGGACGCC TAGAACCTTG AGTGAGTTGC TGAAGCCTTT TGTTTGGGAT AAAAAAATAG  
1860 GTGGGGTAAC GACAAGACAA AAAATTCTTG ACCCTGAGCG TAATCTCGTG ACAATGTTTG  
1920 CTAACCTTGT AGAGGAAATT AGGGCAGAAG GAACTATGAA AGCAATGAGT GTGACTGGTA  
1980 AAGTAGGGTG CTTACCTGGT CGAACAAATTG CTTTTAGAAA TATAGTGGAG ACAGTGTATA

10

20

-159-

CAAAGTTTAT AAAGAGACTT TCATGGGATT TCATAAGGAA GTTCTGATG ATAGAACGTC 2040  
TTACAAATTT GACTTTAAAA AAAGGCTATA AAACGTGTTAT GCAGGATACT TCTGTTGTGT 2100  
5 ATACAGATGC TCCTACAAAGT TGGAAAAAGT TCATTAGACA GCAACTAAGG TGGGCAGAAG 2160  
GTTCTCAGTA TAACAATCTA AAGATGACTC CTTGGGATGAT TAGAAATGCC CCTCTTATGT 2220  
TTTTTATTTA TTTTACAGAT ATGATTTTAC CTATGCTACT TATTAGCTTT GGTGTGAATA 2280  
10 TATTCCTGTT GAAAAATATTA AATATAACTA CAATTGTTTA TACAGCTTCA TGGTGGGAAA 2340  
TTATTTTATA TGTCTTTTG GGAATGATTT TTAGCTTTGG AGGAAGAAAC TTATAAGCTA 2400  
15 TGTCIAGAAT GAAGTGGTAT TATGTATTTT TATTTCCTGT TTTTATAATC GTTTTGAGTA 2460  
TAATTATGTG CCTATTAGG CTATTAGGAC TTATGAGATG TTCCTGATGAT TTAGGGTGGG 2520  
GAACTAGGAA TTAAACAGAG TGAGATAAAT AGTAGTGCGT ATATAGAGTA TTACTCAGA 2580  
20 GTATTAATTG ATTTTGGAA AGGAAAAGTG TTTTTTAATG TTAAGAAAGA ACTTGAATA 2640  
TCAGATTATG ACACGAGCTG GAACAAATTT AGCTATCTTG TTTTTCATTA TATTAGGGAT 2700

-160-

2760 TATTGTTGAA GTTTTGTTTT AAGGCTCATT GTCATCTGTA GTGGCTCACT TCAGACGTAA  
2820 GGGTCATATT TTAATGTGAA AAGAGTGTTA AAAGATTAAAT CACTTATATT ATTTAATAG  
2880 5 AAATAGTGTA AGGAATTGTT ATGAAAAAAG TAAAAAAGC TGTATTCCCT GCTGCAGGGC  
2940 TGGGCACACG ATTTTTCCTT GCCACTAAAG CTTTGGCAAA AGAAATGCTT CCAATTGTAG  
3000 ACCGCCCCAC AATTCATTTT GTCATTGAAG AAGCTTTACG TTCGGGTATT GAAGATAATC  
3060 TAGTAGTTAC TCGAAAGAGT AAACGTTCTA TTGAAGATTA TTTTGATTCA ACTTTTGAAT  
3120 TCGAATATAG TCTTAGAAAA CAAGGAAAGA TGGAACTTCT TAAGTCAGTT AACGAATCGA  
3180 15 CTGATATTAA AGTACATTTC GTTCGTCAA GTTCACCACG TGGTCTTGGT GACGCTGTTC  
3240 TCCAAGCGAA GTCTTTTGTT GGTGACGATC CCTTTGTTGT AATGCTTGGT GATGACCCTTA  
3300 TGGATATCAC CGACTCAACT GCTGTACCTT TAACAAGACA ATTGATGGAT GATTACAACG  
3360 CAACACAGGC TTCAACTATC GCAGTAATGC CTGTTAGATA TGAAGATGTT TCTTCTTATG  
3420 20 GTGTGATTTC TCCTAGATTG GAAAGTAGTA ATGGCTCTTA TAGTGTTGAT GCTTTTGTAG

-161-

AGAAACCAAA ACCAGAAGAA GCGCCTAGCC ATTTAGCTAT TATTGGACGT TATCTACTTA 3480  
CTCCTGAGAT TTTTCTCTATA TTAGAAACCC AAAAGCCAGG AGCAGGTAAT GAAATTCAAT 3540  
5 TGACAGATGC TATTGATACA TTGAATAAGA CACAGAGTGT TTTTGCGCGT GAATTGTGG 3600  
GCAACGTTA CGATGTTGGT GATAAGTTTA ATTTATGAA AACATCAATT GATTATGCTC 3660  
TTCAACATCC TCAGATTAAA GAGAGTTTAA AAAATTACGT TATTGCACCT GGTAAGCAAT 3720  
10 TGGAGAAGCT AGATGACTGT TCGTCAAGTG GACACCTATG AATTGTATAG AAAGTTATCA 3780  
AAAAATGGCTA AATGTCCCTG ATCTTCCAGC TTATTTAAAA GATGAATTGC TCAGCATGGA 3840  
15 TGACAAAAACA AAAGAAGACG CCTTTTACAC AAACCTTGAA TTCGGAACAG CTGGTATGCG 3900  
TGGTTATATT GGTGCTGGGA CAAACCGTAT TAATATCTAT GTGGTGCGTC AAGCACACAG 3960  
AAGGCCTTGC AAATTAGTTG AATCAAAAAG CGAAACCGCC AAAAAAGCTG GGGTTGCTAT 4020  
20 TGCCTATGAC TCGCGACATT TTTTACCAGA ATTTCGCTTTT GAATCTGCCC AAGTATTAGC 4080  
GGCCCATGGC ATTAAATCTT ATGTTTTTGA AAGCCTACGC CCTACTCCTG AGCTGTCTTT 4140

-162-

4200 TGCTGTTCGT CATCTCGGAG CATTTGCTGG TATTATGGTA ACCGCCAGTC ATACCCCTGC  
4260 TCCTTTTAAAT GGTATATAAG TTTACGGTTC TGATGGTGGG CAAATGCTTC CAGCTGACTA  
4320 5 TATTCGTGCG ATTGATAACC CATTTGCTGT AGTCCTTGCT GACTTAGAAG AAGCTAAATC  
4380 AACTGGTCTT ATTGAAGTAA TTGGTGAAAC TCTCGATGCT GCCTACCTTG AAGAGGTAA  
4440 AAGCGTTAAT ATCAATCAAG ATTTGATTGA CCAATACGGT CGCGATATGC AAATTGTCTA  
4500 CACACCTCTT CATGGTACTG GAGAAATGCT AGCACGTGCA GCCTTAGCAC AAGCTGGTTT  
4560 CGAATCTGTT CAAGTTGTG AAGCTCAAGC AAAACCAGAC CCAGACTTCT CAACAGTTGC  
4620 15 ATCACCAAAC CCTGAAAGTC AAGCCGCCCTT TGCCTTAGCT GAAGAACTAG GCGGTCAAGT  
4680 CGATGCTGAT GTATTAGTGG CCACTGACCC TGATGCTGAC CGTCTCGGTG TTGAAATTCG  
4740 TCAAGCTGAT GGCAGTTATT GGAACCTTTC TGGTAACCAA ATCGGTGCTC TTATCGCCAA  
4800 20 ATACATTTTA GAAGCTCACA AACAAAGCTGG GACACTCCCA AAGAATGCTG CATTGGCAAA  
4860 ATCAATAGTA TCAACTGAAT TAGTCACCAA AATTGCGGAA AGCTATGGCG AACCATGTTT



-163-

AACGTCATTA CAGGTTTCAA ATTCAATCGCT GAGAAAATTC AAGAAATTGA AGAAAAACAT 4920

AACCTACATG TTTGGGTTTG AAGAAAGCTG A 4951

5

(2) INFORMATION FOR SEQ ID NO:6:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1307 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "DNA"

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TGAGTTTATT GATTGCCTCC AGCTTGGAGT TAGAATAGGG CATCTGGATG GCATTGTGCA 60

CGTATTTTCT GTAGCGCACC AGCGTGCTAA GAGCAGTTCT AAAGGCTTGA TTGAGTTGGG 120

GAAAAGCCTC AGTTAACAAG TCAAAGAAAT GGTCGGCATT CTTTCTTTC AGGTGGAAAA 180

20

-164-

GCAGGAGCTG GTAAAGATCG TCATAAATAGT GGAGCTCATC TGAGAAAGCT AAGGTTTAT 240  
TGACGATTTC TCGAGGTGTC AGTGCTGTC TAAAAGTTCT TGAGTAGAAG GCCTTATCAG 300  
5 ACAGTTTTCG GCTATCCTTT TGGAAAATTC GCCAGTGATT TTTCATGGCG CGATAGGAAA 360  
GTGATTGCTT GTCAAAATTC TTCATGATGA CAATTCTGGT TGTCAATCATC TCACGGCTTA 420  
GGGTGCTGGA TGATATGAAA TCTATCAAGG ACGATTGTG CATTGGGAAA TAGGCATTTA 480  
10 ATCACAAGAC TCTATTACGT ATCTAGTTGG TCAAATATTC GACCGTCAGT CCTATAAATA 540  
TACATCTAAG ACCAGAGAAG AACAAAAAAC ATCTACGAAA AAGGCTCTCT TAAACAAGGA 600  
TTTCCGTCAA GCTATTGCAT TTGGATTGGA CCGTTATGCC TCTCAGTTGA ATGGACAAAC 660  
15 TGGAGCAAGC AAAATCTTAC GTAATCTCTT TGTCCCACCA ACATTGTTC AAGCAGATGG 720  
CAAAAACTTT GCGGATATGG CCAAAGAGAA ATTGGTCACT TATGGGGATG AATGGAAGGA 780  
20 TGTTAATCTT GCAGATTCTC AGGATGGTCT TTACAATCCA GAAAAAGCAC GGGCTGAATT 840  
TCGTAAAGCT AAATTAGCCT TACAAGCAGA AGGAGTCCAA TTCCCAATTC ATTTAGATAT 900

-165-

GCCAGTTGAC CAGACAGCAA CTACAAAAGT TCAGCGGCTC CAATCTATGA AACATCCTT 960  
GGAAGTAACT TTAGGAGCTG ATAATGTCTAT TATTGATATC CAACAACCTAC AAAAAGACGA 1020  
5 AGTAAACAAT ATTACATATT TTGCTGAAAA TGCTGCTGGC GAAGACTGGG ATTTATCAGA 1080  
TAATGTGCGT TGGGTCCAG ACTTTGCCGA TCCATCAACC TACCTTGATA TCATCAAACC 1140  
ATCTGTAGGA GAAAGTACTA AAACATATTT AGGTTTGAC TCAGGGGAAG ATAATGTAGC 1200  
10 TGCTAAAAAA GTAGGTCTAT ATGACTACGA AAAATTGGTT ACTGAGGCTG GTGATGAGGC 1260  
TACAGATGTT CGTAAACGCT ATGATAAATA CGCTGCAGCC CAAGCTT 1307

15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

-166-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5 Ala Arg Tyr Phe Leu Glu Arg Asp Leu Val His Val Val Ala Ser Asp  
1 5 10 15

Met His Asn Leu Asp Ser Arg Pro Pro Tyr Met Gln Gln Ala Tyr Asp  
20 25 30

10

Ile Ile Ala Lys Lys Tyr Arg Ala Lys Lys Ala Lys Glu Leu Phe Val  
35 40 45

Asp Asn Pro Arg Lys Ile Ile Met Asp His  
50 55

15

(2) INFORMATION FOR SEQ ID NO:8:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

-167-

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Leu Gly Glu Asn Met Lys Glu Gln Asn Thr Leu Glu Ile Asp Val Leu  
1 5 10 15

Gln Tyr Ser Glu Leu Ile Gly Arg Ser Val Ile Leu Leu Val Ala Leu  
20 25 30

Tyr Phe Phe Ser Cys Phe Phe Leu  
35 40

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 87 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

-168-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

5      Leu Pro Tyr Ala Ala Phe Leu Gln Asp Pro Ile Gly Trp Leu Phe Asp  
       1            5            10            15

10     Arg Val Ala Ala Gln Lys Ile Ile Ser Ile Thr Arg Ala Asp Val Ala  
       20            25            30

      His Trp Arg Ser Lys Thr Ala Asp Ile Thr Ala Ser Pro Asn Asn Lys  
       35            40            45

15     Arg Asn Thr Leu Ile Gly Phe Leu Ala Phe Phe Ile Gly Thr Ser Val  
       50            55            60

      Ile Val Leu Leu Leu Glu Leu Leu Asp Thr His Val Lys Arg Pro Glu  
       65            70            75            80

20     Asp Ile Glu Asp Thr Leu Gln  
       85

-169-

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gln His Phe Ser Ser Gly Thr Ala Asp Leu Ser His Gly Leu Cys Asp  
 1 5 10 15  
 Thr Asn Ile Glu Asn Leu Phe Val Val Gln Ser Gly Ser Val Ser Pro  
 20 25 30  
 Asn Pro Thr Ala Leu Ser Gln Ser Lys Asn Phe  
 35 40

(2) INFORMATION FOR SEQ ID NO:11:

-170-

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 394 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

10

Met Lys Ile Ala Ile Ala Gly Ser Gly Tyr Val Gly Leu Ser Leu Ala  
 1 5 10 15

Val Leu Leu Ala Gln His His Glu Val Lys Val Ile Asp Val Ile Lys  
 20 25 30

15

Asp Lys Val Glu Ser Ile Asn Asn Arg Lys Ser Pro Ile Lys Asp Glu  
 35 40 45

Ala Ile Glu Lys Tyr Leu Val Glu Lys Glu Leu Asn Leu Glu Ala Ser  
 50 55 60

20



-171-

Leu Asp Pro Ala His Val Tyr Lys Asp Val Glu Tyr Ala Ile Ile Ala  
65 70 75 80

Thr Pro Thr Asn Tyr Asp Val Asp Leu Asn Gln Phe Asp Thr Ser Ser  
85 90 95

Val Glu Ala Ala Ile Lys Thr Cys Met Glu Tyr Asn Asp Thr Cys Thr  
100 105 110

Ile Val Ile Lys Ser Thr Ile Pro Glu Gly Tyr Thr Lys Glu Val Arg  
115 120 125

Glu Lys Phe Asn Thr Asp Arg Ile Ile Phe Ser Pro Glu Phe Leu Arg  
130 135 140

Glu Ser Lys Ala Leu Tyr Asp Asn Leu Tyr Pro Ser Arg Ile Val Val  
145 150 155 160

Gly Thr Asp Leu Asp Asp Ser Glu Leu Thr Lys Arg Ala Trp Gln Phe  
165 170 175

Ala Asp Leu Leu Lys Gly Gly Ala Ile Lys Glu Glu Val Pro Ile Leu  
180 185 190

5

10

15

20

-172-

Val Val Ala Phe Asn Glu Ala Glu Val Ala Lys Leu Phe Ser Asn Thr  
 195 200 205  
 Tyr Leu Ala Thr Arg Val Ala Tyr Phe Asn Glu Ile Asp Thr Tyr Ser  
 210 215 220  
 Glu Val Lys Gly Leu Asn Pro Lys Thr Ile Ile Asp Ile Val Cys Tyr  
 225 230 235 240  
 Asp Pro Arg Ile Gly Ser Asp Tyr Asn Asn Pro Ser Phe Gly Tyr Gly  
 245 250 255  
 Gly Tyr Cys Leu Pro Lys Asp Thr Lys Gln Leu Lys Ala Ser Phe Arg  
 260 265 270  
 Asp Val Pro Glu Asn Leu Ile Thr Ala Val Val Gln Ser Asn Lys Thr  
 275 280 285  
 Arg Lys Asp Tyr Ile Ala Gly Ala Ile Leu Ala Lys Gln Pro Ser Val  
 290 295 300  
 Val Gly Ile Tyr Arg Leu Ile Met Lys Ser Asp Ser Asp Asn Phe Arg  
 305 310 315 320

-173-

Ser Ser Ala Val Lys Gly Val Met Glu Arg Leu Asp Asn Tyr Gly Lys  
 325 330 335

Glu Ile Val Ile Tyr Glu Pro Thr Ile Glu Cys Asp Thr Phe Met Gly  
 340 345 350

Tyr Arg Val Ile Lys Ser Leu Asp Glu Phe Lys Asn Ile Ser Asp Ile  
 355 360 365

Val Val Ala Asn Arg Met Asn Asp Asp Leu Arg Asp Ile Gln Glu Lys  
 370 375 380

Leu Tyr Thr Arg Asp Leu Phe Gly Arg Glu  
 385 390

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 416 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

15

-174-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

5	Met Tyr Thr Phe Ile Leu Met Leu Leu Asp Phe Phe Gln Asn His Asp	15
	1 5 10	
	Phe His Phe Phe Met Leu Phe Phe Val Phe Ile Leu Ile Arg Trp Ala	30
	20 25	
10	Val Ile Tyr Phe His Ala Val Arg Tyr Lys Ser Tyr Ser Cys Ser Val	45
	35 40	
	Ser Asp Glu Lys Leu Phe Ser Ser Val Ile Ile Pro Val Val Asp Glu	60
	50 55	
15	Pro Leu Asn Leu Phe Glu Ser Val Leu Asn Arg Ile Ser Arg His Lys	80
	65 70 75	
20	Pro Ser Glu Ile Ile Val Val Ile Asn Gly Pro Lys Asn Glu Arg Leu	95
	85 90	

-175-

Val Lys Leu Cys His Asp Phe Asn Glu Lys Leu Glu Asn Asn Met Thr  
 100 105 110  
 Pro Ile Gln Cys Tyr Thr Pro Val Pro Gly Lys Arg Asn Ala Ile  
 115 120 125  
 Arg Val Gly Leu Glu His Val Asp Ser Gln Ser Asp Ile Thr Val Leu  
 130 135 140  
 Val Asp Ser Asp Thr Val Trp Thr Pro Arg Thr Leu Ser Glu Leu Leu  
 145 150 155 160  
 Lys Pro Phe Val Cys Asp Lys Lys Ile Gly Gly Val Thr Thr Arg Gln  
 165 170 175  
 Lys Ile Leu Asp Pro Glu Arg Asn Leu Val Thr Met Phe Ala Asn Leu  
 180 185 190  
 Leu Glu Glu Ile Arg Ala Glu Gly Thr Met Lys Ala Met Ser Val Thr  
 195 200 205  
 Gly Lys Val Gly Cys Leu Pro Gly Arg Thr Ile Ala Phe Arg Asn Ile  
 210 215 220

-176-

Val Glu Arg Val Tyr Thr Lys Phe Ile Glu Glu Thr Phe Met Gly Phe  
 225 230 235 240

His Lys Glu Val Ser Asp Asp Arg Ser Leu Thr Asn Leu Thr Leu Lys  
 245 250 255

Lys Gly Tyr Lys Thr Val Met Gln Asp Thr Ser Val Val Tyr Thr Asp  
 260 265 270

Ala Pro Thr Ser Trp Lys Lys Phe Ile Arg Gln Gln Leu Arg Trp Ala  
 275 280 285

Glu Gly Ser Gln Tyr Asn Asn Leu Lys Met Thr Pro Trp Met Ile Arg  
 290 295 300

Asn Ala Pro Leu Met Phe Phe Ile Tyr Phe Thr Asp Met Ile Leu Pro  
 305 310 315 320

Met Leu Leu Ile Ser Phe Gly Val Asn Ile Phe Leu Leu Lys Ile Leu  
 325 330 335

Asn Ile Thr Thr Ile Val Tyr Thr Ala Ser Trp Trp Glu Ile Ile Leu  
 340 345 350

5

10

15

20

-177-

Tyr Val Leu Leu Gly Met Ile Phe Ser Phe Gly Gly Arg Asn Phe Lys  
 355 360 365  
 Ala Met Ser Arg Met Lys Trp Tyr Tyr Val Phe Leu Ile Pro Val Phe  
 370 375 380  
 Ile Ile Val Leu Ser Ile Ile Met Cys Pro Ile Arg Leu Leu Gly Leu  
 385 390 395 400  
 Met Arg Cys Ser Asp Asp Leu Gly Trp Gly Thr Arg Asn Leu Thr Glu  
 405 410 415

## (2) INFORMATION FOR SEQ ID NO:13:

15

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 306 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

## (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

5

10

15

20

Lys Val His Phe Val Arg Gln Ser Ser Pro Arg Gly Leu Gly Asp Ala  
100 105 110



-179-

Val Leu Gln Ala Lys Ser Phe Val Gly Asp Asp Pro Phe Val Val Met  
 115 120 125  
 Leu Gly Asp Asp Leu Met Asp Ile Thr Asp Ser Thr Ala Val Pro Leu  
 130 135 140  
 Thr Arg Gln Leu Met Asp Asp Tyr Asn Ala Thr Gln Ala Ser Thr Ile  
 145 150 155 160  
 Ala Val Met Pro Val Arg Tyr Glu Asp Val Ser Ser Tyr Gly Val Ile  
 165 170 175  
 Ser Pro Arg Leu Glu Ser Ser Asn Gly Leu Tyr Ser Val Asp Ala Phe  
 180 185 190  
 Val Glu Lys Pro Lys Pro Glu Glu Ala Pro Ser His Leu Ala Ile Ile  
 195 200 205  
 Gly Arg Tyr Leu Leu Thr Pro Glu Ile Phe Ser Ile Leu Glu Thr Gln  
 210 215 220  
 Lys Pro Gly Ala Gly Asn Glu Ile Gln Leu Thr Asp Ala Ile Asp Thr  
 225 230 235 240

-180-

Leu Asn Lys Thr Gln Ser Val Phe Ala Arg Glu Phe Val Gly Lys Arg  
245 250 255

Tyr Asp Val Gly Asp Lys Phe Asn Phe Met Lys Thr Ser Ile Asp Tyr  
260 265 270

5

Ala Leu Gln His Pro Gln Ile Lys Glu Ser Leu Lys Asn Tyr Val Ile  
275 280 285

Ala Leu Gly Lys Gln Leu Glu Lys Leu Asp Asp Cys Ser Ser Ser Gly  
290 295 300

10

His Leu  
305

15

## (2) INFORMATION FOR SEQ ID NO:14:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 397 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

-181-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

5	Met Asn Cys Ile Glu Ser Tyr Gln Lys Trp Leu Asn Val Pro Asp Leu	15
1	5	10
10	Pro Ala Tyr Leu Lys Asp Glu Leu Leu Ser Met Asp Asp Lys Thr Lys	30
	20	25
	Glu Asp Ala Phe Tyr Thr Asn Leu Glu Phe Gly Thr Ala Gly Met Arg	45
	35	40
15	Gly Tyr Ile Gly Ala Gly Thr Asn Arg Ile Asn Ile Tyr Val Val Arg	60
	50	55
	Gln Ala His Arg Arg Pro Cys Lys Leu Val Glu Ser Lys Gly Glu Thr	80
	65	70
20	Ala Lys Lys Ala Gly Val Ala Ile Ala Tyr Asp Ser Arg His Phe Ser	95
	85	90

-182-

Pro Glu Phe Ala Phe Glu Ser Ala Gln Val Leu Ala Ala His Gly Ile  
 100 105 110

Lys Ser Tyr Val Phe Glu Ser Leu Arg Pro Thr Pro Glu Leu Ser Phe  
 115 120 125

Ala Val Arg His Leu Gly Ala Phe Ala Gly Ile Met Val Thr Ala Ser  
 130 135 140

His Thr Pro Ala Pro Phe Asn Gly Tyr Lys Val Tyr Gly Ser Asp Gly  
 145 150 155 160

Gly Gln Met Leu Pro Ala Asp Tyr Ile Arg Ala Ile Asp Asn Pro Phe  
 165 170 175

Ala Val Val Leu Ala Asp Leu Glu Glu Ala Lys Ser Thr Gly Leu Ile  
 180 185 190

Glu Val Ile Gly Glu Thr Leu Asp Ala Ala Tyr Leu Glu Glu Val Lys  
 195 200 205

Ser Val Asn Ile Asn Gln Asp Leu Ile Asp Gln Tyr Gly Arg Asp Met  
 210 215 220

-183-

Gln Ile Val Tyr Thr Pro Leu His Gly Thr Gly Glu Met Leu Ala Arg  
 225 230 235 240

Arg Ala Leu Ala Gln Ala Gly Phe Glu Ser Val Gln Val Val Glu Ala  
 245 250 255

Gln Ala Lys Pro Asp Pro Asp Phe Ser Thr Thr Val Ala Ser Pro Asn Pro  
 260 265 270

Glu Ser Gln Ala Ala Phe Ala Leu Ala Glu Glu Leu Gly Arg Gln Val  
 275 280 285

Asp Ala Asp Val Leu Val Ala Thr Asp Pro Asp Ala Asp Arg Leu Gly  
 290 295 300

Val Glu Ile Arg Gln Ala Asp Gly Ser Tyr Trp Asn Leu Ser Gly Asn  
 305 310 315 320

Gln Ile Gly Ala Leu Ile Ala Lys Tyr Ile Leu Glu Ala His Lys Gln  
 325 330 335

Ala Gly Thr Leu Pro Lys Asn Ala Ala Leu Ala Lys Ser Ile Val Ser  
 340 345 350

Thr Glu Leu Val Thr Lys Ile Ala Glu Ser Tyr Gly Glu Pro Cys Leu  
355 360 365

Thr Ser Leu Gln Val Ser Asn Ser Ser Leu Arg Lys Phe Lys Asn Leu  
370 375 380

Lys Lys Asn Ile Thr Tyr Met Phe Gly Phe Glu Glu Ser  
385 390 395

10

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 274 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: protein

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

-185-

Gln Asp Ser Ile Thr Tyr Leu Val Gly Gln Ile Phe Asp Arg Gln Ser  
 1 5 10 15

Tyr Lys Tyr Thr Ser Lys Thr Arg Glu Glu Gln Lys Thr Ser Thr Lys  
 20 25 30

Lys Ala Leu Leu Asn Lys Asp Phe Arg Gln Ala Ile Ala Phe Gly Phe  
 35 40 45

Asp Arg Tyr Ala Ser Gln Leu Asn Gly Gln Thr Gly Ala Ser Lys Ile  
 50 55 60

Leu Arg Asn Leu Phe Val Pro Pro Thr Phe Val Gln Ala Asp Gly Lys  
 65 70 75 80

Asn Phe Gly Asp Met Ala Lys Glu Lys Leu Val Thr Tyr Gly Asp Glu  
 85 90 95

Trp Lys Asp Val Asn Leu Ala Asp Ser Gln Asp Gly Leu Tyr Asn Pro  
 100 105 110

Glu Lys Ala Arg Ala Glu Phe Arg Lys Ala Lys Leu Ala Leu Gln Ala  
 115 120 125

-186-

Glu Gly Val Gln Phe Pro Ile His Leu Asp Met Pro Val Asp Gln Thr  
 130 135 140  
 Ala Thr Thr Lys Val Gln Arg Val Gln Ser Met Lys Gln Ser Leu Glu  
 145 150 155 160  
 Val Thr Leu Gly Ala Asp Asn Val Ile Ile Asp Ile Gln Gln Leu Gln  
 165 170 175  
 Lys Asp Glu Val Asn Asn Ile Thr Tyr Phe Ala Glu Asn Ala Ala Gly  
 180 185 190  
 Glu Asp Trp Asp Leu Ser Asp Asn Val Gly Trp Gly Pro Asp Phe Ala  
 195 200 205  
 Asp Pro Ser Thr Tyr Leu Asp Ile Ile Lys Pro Ser Val Gly Glu Ser  
 210 215 220  
 Thr Lys Thr Tyr Leu Gly Phe Asp Ser Gly Glu Asp Asn Val Ala Ala  
 225 230 235 240  
 Lys Lys Val Gly Leu Tyr Asp Tyr Glu Lys Leu Val Thr Glu Ala Gly  
 245 250 255



-187-

Asp Glu Ala Thr Asp Val Arg Lys Arg Tyr Asp Lys Tyr Ala Ala Ala  
260 265 270

Gln Ala

5

(2) INFORMATION FOR SEQ ID NO:16:

10 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 177 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

20 Leu Asn Ala Tyr Phe Gln Met His Lys Ser Ser Leu Ile Asp Phe Ile  
1 5 10 15

-188-

Ser Ser Ser Thr Leu Ser Arg Glu Met Met Thr Thr Arg Ile Val Ile  
 20 25 30

Met Lys Asn Phe Asp Lys Gln Ser Leu Ser Tyr Arg Ala Met Lys Asn  
 35 40 45

His Trp Arg Ile Phe Gln Lys Asp Ser Arg Lys Leu Ser Asp Lys Ala  
 50 55 60

Phe Tyr Ser Arg Thr Phe Arg Gln Thr Leu Thr Pro Arg Glu Ile Val  
 65 70 75 80

Asn Lys Thr Leu Ala Phe Ser Asp Glu Leu His Tyr Tyr Asp Asp Leu  
 85 90 95

Tyr Gln Leu Leu Leu Phe His Leu Gln Glu Lys Asn Ala Asp His Phe  
 100 105 110

Phe Asp Leu Leu Thr Glu Ala Phe Pro Gln Leu Asn Gln Ala Phe Arg  
 115 120 125

Thr Ala Leu Ser Thr Leu Val Arg Tyr Arg Lys Tyr Val Thr Asn Ala  
 130 135 140

-189-

Ile Gln Met Pro Tyr Ser Asn Ser Lys Leu Glu Ala Ile Asn Lys Leu  
145 150 155 160

Ile Ser Phe Leu Gln Thr Gln Thr Cys Arg Leu Cys Phe Ser Ser Asn  
165 170 175

5

Ser

10 (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "DNA"

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCCACTATCG ACTACGCG

18

- 190 -

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- 10 (A) DESCRIPTION: /desc = "DNA"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

15 TCATTTGATA TGCCTCCG 18

## (2) INFORMATION FOR SEQ ID NO:19:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## 25 (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "DNA"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

30 GTGAGATAAA TAGTAGTGCG 20

## (2) INFORMATION FOR SEQ ID NO:20:

## 35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs

- 191 -

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5       (ii) MOLECULE TYPE: other nucleic acid  
          (A) DESCRIPTION: /desc = "DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

10   TCCAGCTCGT GTCATAATCT

20

- 192 -

CLAIMS

1. A nucleic acid segment less than about 10 kb in  
5 length that comprises a non-type specific *S. pneumoniae*  
*cps* gene flanking region of sufficient length to allow  
hybridization under standard hybridization conditions to  
a *S. pneumoniae cps* gene flanking region.
- 10 2. The nucleic acid segment of claim 1, wherein the  
segment comprises a non-type specific *S. pneumoniae cps*  
gene 5' flanking region.
- 15 3. The nucleic acid segment of claim 2, wherein the  
segment includes a non-type specific *S. pneumoniae cps*  
gene 5' flanking region encoding for a peptide comprising  
SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 or SEQ ID NO:10.
- 20 4. The nucleic acid segment of claim 3, wherein the  
segment includes a non-type specific *S. pneumoniae cps*  
gene 5' flanking region encoding for a peptide comprising  
25 SEQ ID NO:7.
5. The nucleic acid segment of claim 3, wherein the  
segment includes a non-type specific *S. pneumoniae cps*  
30 gene 5' flanking region encoding for a peptide comprising  
SEQ ID NO:8.
6. The nucleic acid segment of claim 3, wherein the  
35 segment includes a non-type specific *S. pneumoniae cps*

- 193 -

gene 5' flanking region encoding for a peptide comprising  
SEQ ID NO:9.

7. The nucleic acid segment of claim 3, wherein the  
5 segment includes a non-type specific *S. pneumoniae cps*  
gene 5' flanking region encoding for a peptide comprising  
SEQ ID NO:10.

10 8. The nucleic acid segment of claim 2, wherein the  
segment comprises a non-type specific *S. pneumoniae cps*  
gene 5' flanking region having a sequence that  
corresponds to at least a 60 nucleotide long contiguous  
stretch of SEQ ID NO:4.

15

9. The nucleic acid segment of claim 8, wherein the  
segment comprises a non-type specific *S. pneumoniae cps*  
gene 5' flanking region having a sequence that  
20 corresponds to at least a 100 nucleotide long contiguous  
stretch of SEQ ID NO:4.

10. The nucleic acid segment of claim 9, wherein the  
25 segment comprises a non-type specific *S. pneumoniae cps*  
gene 5' flanking region having a sequence that  
corresponds to at least a 500 nucleotide long contiguous  
stretch of SEQ ID NO:4.

30

11. The nucleic acid segment of claim 10, wherein the  
segment comprises a non-type specific *S. pneumoniae cps*  
gene 5' flanking region having a sequence that  
corresponds to SEQ ID NO:4.

35

- 194 -

12. The nucleic acid segment of claim 2, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 3' flanking region.

5

13. The nucleic acid segment of claim 12, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 3' flanking region having a sequence that corresponds to at least a 15 nucleotide long contiguous stretch of SEQ ID NO:6.

10

14. The nucleic acid segment of claim 13, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 3' flanking region having a sequence that corresponds to at least a 30 nucleotide long contiguous stretch of SEQ ID NO:6.

15

15. The nucleic acid segment of claim 14, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 3' flanking region having a sequence that corresponds to at least a 60 nucleotide long contiguous stretch of SEQ ID NO:4.

20

25

16. The nucleic acid segment of claim 15, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 3' flanking region having a sequence that corresponds to at least a 100 nucleotide long contiguous stretch of SEQ ID NO:6.

30

17. The nucleic acid segment of claim 16, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 3' flanking region having a sequence that

35



- 195 -

corresponds to at least a 500 nucleotide long contiguous stretch of SEQ ID NO:6.

5 18. The nucleic acid segment of claim 17, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 3' flanking region having a sequence that corresponds to SEQ ID NO:6.

10

19. The nucleic acid segment of claim 1, wherein the segment comprises a non-type specific *S. pneumoniae cps* gene 5' flanking region and a non-type specific *S. pneumoniae cps* gene 3' flanking region.

15

20. The nucleic acid segment of claim 19, wherein the segment comprises a 5' flanking region that encodes for a peptide comprising SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 or SEQ ID NO:10 and a 3' flanking region sequence that corresponds to at least a 30 nucleotide long contiguous stretch of SEQ ID NO:6.

25 21. The nucleic acid segment of claim 20, wherein the segment comprises a 5' flanking region that encodes for a peptide comprising SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 or SEQ ID NO:10 and a 3' flanking region sequence that corresponds to at least a 100 nucleotide long contiguous stretch of SEQ ID NO:6.

30

22. The nucleic acid segment of claim 21, wherein the segment comprises a 5' flanking region that encodes for a peptide comprising SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9

35

- 196 -

or SEQ ID NO:10 and a 3' flanking region sequence that corresponds to SEQ ID NO:6.

5     23. The nucleic acid segment of claim 19, wherein the  
segment comprises a 5' flanking region sequence that  
corresponds to at least a 60 nucleotide long contiguous  
stretch of SEQ ID NO:4 and a 3' flanking region sequence  
that corresponds to at least a 30 nucleotide long  
10    contiguous stretch of SEQ ID NO:6.

24. The nucleic acid segment of claim 23, wherein the  
segment comprises a 5' flanking region sequence that  
15    corresponds to at least a 100 nucleotide long contiguous  
stretch of SEQ ID NO:4 and a 3' flanking region sequence  
that corresponds to at least a 100 nucleotide long  
contiguous stretch of SEQ ID NO:6.

20    25. The nucleic acid segment of claim 24, wherein the  
segment comprises a 5' flanking region sequence that  
corresponds to SEQ ID NO:4 and a 3' flanking region  
sequence that corresponds to SEQ ID NO:6.

25    26. The nucleic acid segment of claim 1, further defined  
as including a type specific *S. pneumoniae* *cps* gene  
region of sufficient length to allow hybridization to a  
30    *S. pneumoniae* *cps* gene region under standard  
hybridization conditions.

27. The nucleic acid segment of claim 26, further  
35    defined as less than about 5,000 nucleotides in length.

- 197 -

28. The nucleic acid segment of claim 27, further defined as less than about 1,000 nucleotides in length.

5 29. The nucleic acid segment of claim 1, further defined as a recombinant vector.

10 30. A nucleic acid cassette less than about 20 kb in length that comprises a non-type specific *S. pneumoniae* *cps* gene 5' flanking region sequence and a non-type specific *S. pneumoniae* *cps* gene 3' flanking region sequence, the flanking region sequences being of sufficient length to allow hybridization under standard  
15 hybridization conditions to a *S. pneumoniae* *cps* gene flanking region.

20 31. A nucleic acid segment of up to about 20 kb in length, comprising a *S. pneumoniae* *cps* gene region of sufficient length to allow hybridization to a *S. pneumoniae* *cps* gene region under standard hybridization conditions.

25

32. The nucleic acid segment of claim 31, further defined as comprising a *cpsB* gene.

30 33. The nucleic acid segment of claim 31, further defined as comprising a *cpsC* gene.

35 34. The nucleic acid segment of claim 31, further defined as comprising a *cpsE* gene.

- 198 -

35. The nucleic acid segment of claim 31, further defined as comprising a *cpsD* gene.

5

36. The nucleic acid segment of claim 31, further defined as comprising a *cpsS* gene.

10 37. The nucleic acid segment of claim 31, further defined as comprising a *cpsU* gene.

15 38. The nucleic acid segment of claim 31, further defined as comprising a *cpsM* gene.

20 39. The nucleic acid segment of claim 31, further defined as comprising a *plpA* gene.

40. The nucleic acid segment of claim 31, further defined as comprising a *tnpA* gene.

25

41. The nucleic acid segment of claim 31, further defined as comprising a complete *S. pneumoniae cps* gene region.

30

42. The nucleic acid segment of claim 31, wherein the *S. pneumoniae cps* gene region is defined as a Type 3 *cps* gene region.

35

- 199 -

43. The nucleic acid segment of claim 42, further defined as comprising a Type 3 *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *tnpA* and *'plpA* gene.

5

44. The nucleic acid segment of claim 31, further defined as comprising a *cps* gene flanking region, wherein the flanking region corresponds to any nucleic acid segment in accordance with the foregoing claims 1 through  
10 25.

45. The nucleic acid segment of claim 44, further defined as less than about 10,000 nucleotides in length.  
15

46. The nucleic acid segment of claim 45, further defined as less than about 5,000 nucleotides in length.

20

47. The nucleic acid segment of claim 44, further defined as a DNA cassette bounded at each terminus by a PCR primer of known sequence or a restriction enzyme recognition site.

25

48. The nucleic acid segment of claim 47, wherein the segment is bounded by an *SphI* or *SalI* site.

30

49. The nucleic acid segment of claim 44, further defined as a recombinant vector.

35

50. The nucleic acid segment of claim 49, further defined as recombinant vector comprising at least one

- 200 -

*S. pneumoniae cps* gene and sufficient flanking region to allow homologous recombination of the fragment in a *S. pneumoniae* host cell.

5

51. The nucleic acid segment of claim 50, further defined as comprising a complete *S. pneumoniae cps* gene region.

10

52. A recombinant host cell comprising a recombinant vector comprising a nucleic acid segment in accordance with claim 44.

15

53. The recombinant host cell of claim 52, further defined as a recombinant *E. coli* host cell.

20

54. The recombinant host cell of claim 52, further defined as a recombinant gram positive host cell.

25

55. The recombinant host cell of claim 54, further defined as a *Bacillus*, *Staphylococcus*, or *Streptococcus* host cell.

30

56. The recombinant host cell of claim 55, further defined as a recombinant *S. pneumoniae* host cell.

35

57. A recombinant host cell in accordance with claim 52, further defined as including an engineered resistance gene.

- 201 -

58. A recombinant *S. pneumoniae* cell of a selected serotype, the cell expressing a *cps* gene of another *S. pneumoniae* serotype.

5

59. The recombinant *S. pneumoniae* cell of claim 58, expressing a *cpsB*, *cpsC*, *cpsE*, *cpsD*, *cpsS*, *cpsU*, *cpsM*, *plpA* or *tnpA* gene.

10

60. A method for preparing a recombinant host cell, comprising preparing a *S. pneumoniae cps* gene and transforming a host cell with said gene.

15

61. The method of claim 60, wherein the host cell is defined as a *S. pneumoniae* host cell, and the *cps* gene is introduced by a method comprising the steps of:

20

(a) preparing a DNA segment that includes a selected *S. pneumoniae cps* gene flanked by sufficient *S. pneumoniae* flanking regions to allow homologous recombination in the *S. pneumoniae* host;

25

(b) transforming the *S. pneumoniae* host with the DNA segment; and

30

(c) selecting a recombinant host that expresses the *S. pneumoniae cps* gene.

62. The method of claim 61, wherein the DNA segment is a plasmid.

35

- 202 -

63. The method of claim 61, wherein the host, prior to transformation, is a high producer of the capsular polysaccharides.

5

64. The method of claim 63, wherein corresponding *cps* gene of the host has been replaced by homologous recombination with the recombinant *cps* gene.

10

65. The method of claim 61, wherein the cell is selected by means of a resistance gene.

15

66. The method of claim 65, wherein the resistance gene is positioned in the non type specific *cps* region.

20

67. The method of claim 66, wherein the resistance gene is an erythromycin resistance gene.

25

68. A method for detecting *S. pneumoniae* in a sample, comprising the steps of:

30

(a) obtaining nucleic acids from a sample suspected of containing *S. pneumoniae*;

35

(b) subjecting said nucleic acids to hybridization with a *S. pneumoniae cps* nucleic acid segment comprising a *cps* gene flanking region or a *cps* gene coding region of sufficient length to allow hybridization to *S. pneumoniae cps* nucleic acids under standard hybridization conditions; and



- 203 -

(c) detecting the hybridized nucleic acids.

69. The method of claim 68, wherein said *S. pneumoniae*  
5 *cps* nucleic acid segment comprises a non-type specific  
*S. pneumoniae cps* gene flanking region of sufficient  
length to allow hybridization under standard  
hybridization conditions to a *S. pneumoniae cps* gene  
flanking region.

10

70. The method of claim 68, wherein the nucleic acids  
from said sample are subjected to restriction enzyme  
digestion and size separation prior to hybridization with  
15 said *S. pneumoniae cps* nucleic acid segment.

71. The method of claim 70, wherein the nucleic acids  
are subjected to *SphI* digestion.

20

72. The method of claim 68, wherein said detection of  
hybridized nucleic acid involves PCR.

25

73. A method for determining the capsule type of an  
unknown *S. pneumoniae* strain, comprising obtaining  
nucleic acids from the strain and hybridizing said  
nucleic acids with a *S. pneumoniae cps* DNA segment  
30 comprising either:

(a) a non-type specific *S. pneumoniae cps* gene  
flanking region of sufficient length to  
allow hybridization under standard  
35 hybridization conditions to a

- 204 -

*S. pneumoniae cps* gene flanking region; or

- 5 (b) a type specific *S. pneumoniae cps* gene region of sufficient length to allow hybridization to a *S. pneumoniae cps* gene under standard hybridization conditions.
- 10 74. A method of generating an antibody response, comprising administering to an animal an immunologically effective amount of a Cps peptide or protein.
- 15 75. The method of claim 74, wherein the Cps peptide or protein is encoded for by any one of the nucleic acid sequences in the foregoing claims 32 through 40.
- 20 76. A method for detecting *S. pneumoniae* in a sample, comprising the steps of:
- 25 (a) obtaining proteins from a sample suspected of containing *S. pneumoniae*;
- (b) binding said proteins with an antibody;
- (c) detecting the bound proteins.

- 205 -

77. The method of claim 76, wherein said antibody  
corresponds to an antibody directed against a Cps protein  
5 or peptide.

78. The method of claim 77, wherein said antibody is  
labeled.  
10

79. The method of claim 75, wherein said proteins are  
separated by electrophoresis.

15  
80. A method for preventing infection of a subject with  
*S. pneumoniae* by administering a composition comprising  
an antibody directed against a Cps protein or peptide.

1/17

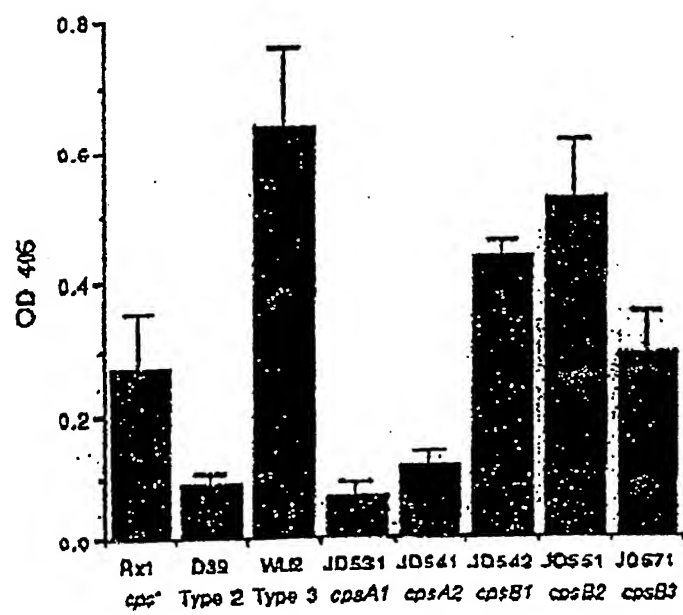


FIG. 1

2/17

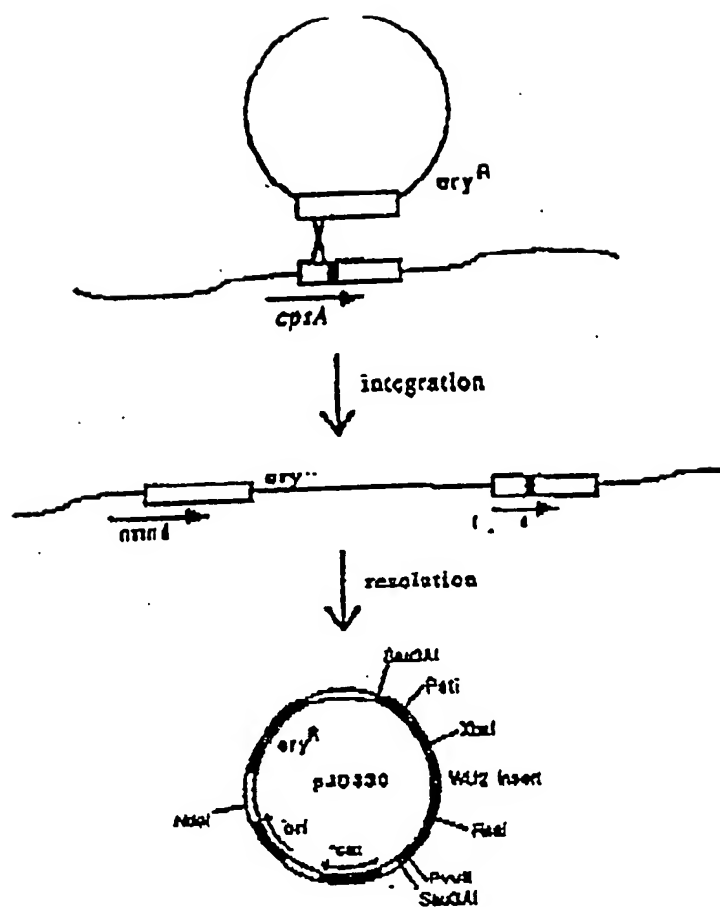


FIG. 2

3/17

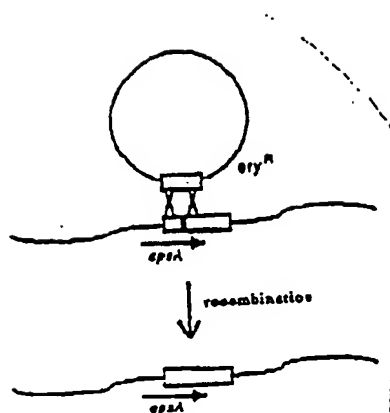


FIG. 3A

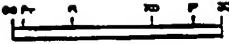
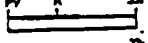





Plasmid		Cps <sup>+</sup> /total CFU	capsule restoration
pJ0330		88/215	+
pJ0337		117/224	+
pJ0361		0/227	-
pJ0343		0/176	-
pJ0345		59/146	+
pJ0369		112/449	+
pJ0380		10/650	+

FIG. 3B

4/17

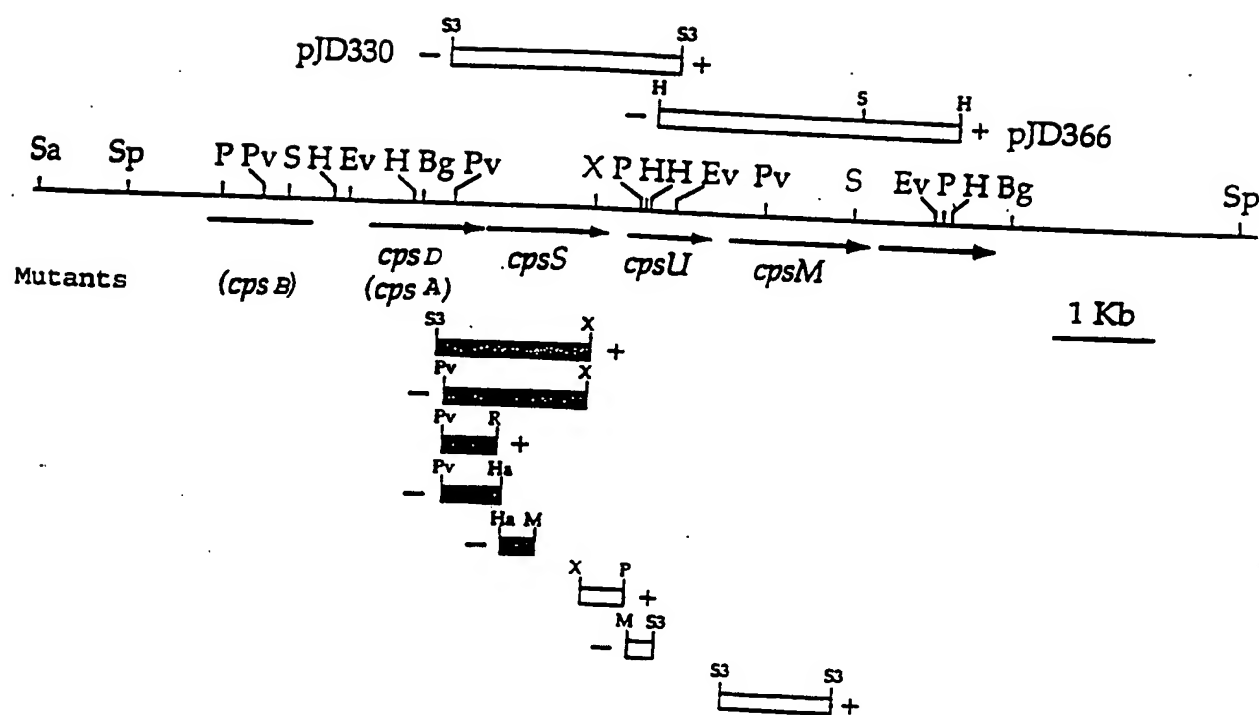


FIG. 4

5/17

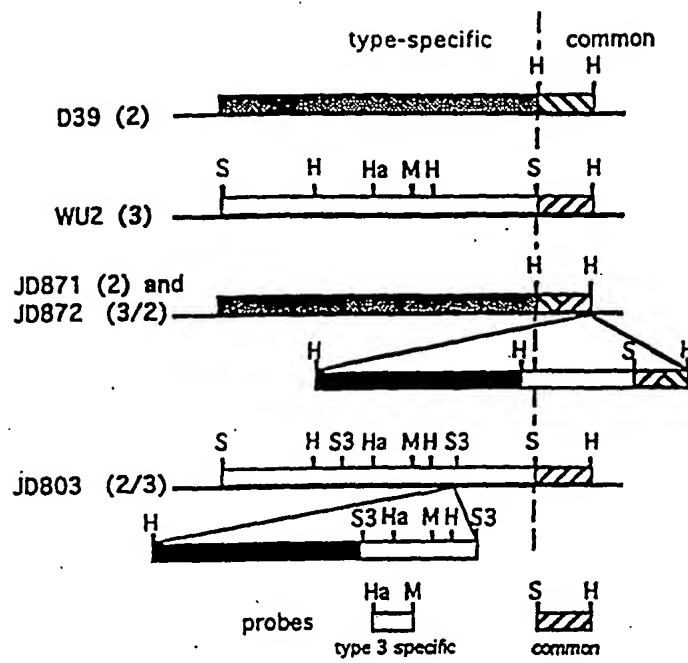


FIG. 5



6/17

SacI 'Cps3B

GAGCT CGG TAT TTT TTG GAA CGT GAT TTA GTT CAT GTA GTT GCA AGT GAC ATG CAC AAT TTA  
ala arg tyr phe leu glu arg asp leu val his val val ala ser asp met his asn leu

GAC AGT AGA CCT CCA TAT ATg CAA CAG GCA TAT gAT ATc ATT GCT AAG AAA TAT AGA GCG  
asp ser arg pro pro tyr met gln gln ala tyr asp ile ile ala lys lys tyr arg ala

AAA AAA GCG AAA GAA CTT TTT GTA GAT AAT CCC AGA AAA ATT ATA ATG GAT CAT TAA TTA  
lys lys ala lys glu leu phe val asp asn pro arg lys ile ile met asp his \*\*\* leu

## start Cps3C

GGA GAA AAT ATG AAG GAA CAA AAC ACT TTG GAA atC GAT GTA TTG CAG TAT Tca gaG CTT  
gly glu asn met lys glu gln asn thr leu glu ile asp val leu gln tyr ser glu leu

ATT GGa aGa aGT GtC att TTA TTA GTG GcA TTA TaC TTC TTC aGT TGC TTT TtC CTa C  
ile gly arg ser val ile leu leu val ala leu tyr phe phe ser cys phe phe leu

-200 nt gap in cps3C

FIG. 6A

cps3C (continued to PstI site. -45 nt expected beyond PstI to Cps3C stop)

ctt ccg tat gcc gcG TTC CTg caa gac ccc atC GGC Tgg CTC TTT Gat aga GTa GCT gcc  
leu pro tyr ala ala phe leu gln asp pro ile gly trp leu phe asp arg val ala ala

caa aaa att atC AGT ATT ACT CGT GCT GAT GTg gca cAC TGG AGG Agc aag acc gcc GAT  
gln lys ile ile ser ile thr arg ala asp val ala his trp arg ser lys thr ala asp

ATC ACC gCT TcG CCA AAT aat AAA cGC AAT ACA CTA ATT GGT TTT TTG Gca TTT TTT ATT  
ile thr ala ser pro asn asn lys arg asn thr leu ile gly phe leu ala phe phe ile

GGA ACT AGT GTc ATA GTT CTT CTT CTT GAA CTT TTG GAC ACT CAT GTG AAA cGT CCG GAA  
gly thr ser val ile val leu leu leu glu leu leu asp thr his val lys arg pro glu

PstI

GAT ATC GAA GAT ACA CTG CAG  
asp ile glu asp thr leu gln

FIG. 6B

-360 nt gap

## 'Cps3E

PvuII

CAG CAC ttc tCA TCT GGC ACA GCT Gat tta TCT CAC Ggc tta TGT Gat ACA AAT ATT gAA  
gln his phe ser ser gly thr ala asp leu ser his gly leu cys asp thr asn ile glu

AAT TTA TTT GTA GTT CAA TCG GGA TCT GTA TCA CCA AAC CCT ACA GCC TTG Tca CAA AGc  
asn leu phe val val gln ser gly ser val ser pro asn pro thr ala leu ser gln ser

End homology

with Cps19fD Repeat

AAA AAT TTT GTG GTT ATG GTA AAG cTT TTT TCA AAA GAG GTC AGT ATA TTG AGT TGG TGG  
lys asn phe

AAA CGA TAA AAT ACA TAA TAT TTT ATT CCT TTG GTT ATC AAA TTA GCC CCT CCT GAA GCT

SacI

CCC CAA TTG ACG GCT TGA GCT C

FIG. 6C

7/17

[illegible][illegible]

AGCGATATACACAGAAACGCTATACACGCCATTTATTTCCACAGCAATAAGCCGAAATAAATTTTATGTATACATCATTTATTTAAATGTCGTGCATTTTTCACAAATCATCATTCAT 2260  
 RD IDQEKLYTDROLFGRG \* \* \* \* \* N T T I L N L L D F F O O N D N F -  
 TTCTTATCTCTCTTTTCTCTCTATCTCTCTCTCTCTATATATTTCTCTCTCTCATATACAGCTCTACAGTTCTACTGTAACTGATACAGAAATTTACTGTCGTAA 2400  
 R F L L F F V F I L I R N A V I T F N A V T R Y K S Y C S V S D E K L F S S V I -  
 ATCTCTCTCTGCGATCAACCACTTAATCTTTTCAAGCTCTACTCAATACAAATTTCCACAGACATAAACCTCGGAATTATGTGCTTATTAACTGCGCAAAAACGACAGCTGTAAAA 2520  
 I P V V D E F L N L F E S V L R N E P S E C I V I V I N C N E R L R -  
 CTTTCTCATGATTTTATCAAAAATTAAGAAATAATCACTCACTCAATGTTATTACCTCTCTCTCCACAGAAATCGTATCGGCTCTCGCAACATGCGCATTCGCAC 2640  
 L C N D F M E K E L E R N H T P I Q C Y T T F V P G E R N A I R V G L E N V D S O -  
 AGTGATATACACTTCTACTGATAGTATACAGTATCGACCCCTCAAGCACTCTACTGCTTCTCAAGCTTTCTTTGCCATAAAAATACTGTCGGGTAAACCAACAGAAAAT 2760  
 S D I T V L V D S D I T M P T R L S E L L R F V C D K E I C G V T T A O K I -  
 CTTGACCTCAGCGTALVCTGTGCAGATTTCTTCAACTCTCTACAGCAATAAGCCGACAGTAATCAAGCAATCACTGTCACTGTAAATACGGCTCTTACTGCTCGAAC 2880  
 L D P E R N L V T H F A N L L E E I R A E G T M K A N S V T C K V C C L P G R T -  
 ATTCTTTTACAAATATAGTCCACAGAGTGTATACAAAGTTATACAGACAGACTTTCATCGCAATTTCATAACCACTTTCTCATATACAGACTTTACAAATTCCTATAAAAACCG 3000  
 A J F R N H I E R V E R Y T R F I E E T F H G F N E E V S D D R E L T S L T L E K E  
 TATAAAATCTCTACAGCATCTCTCTGTATACAGACAGCTCTCTACAGCTCGCAAAAGTCTATACAGACAGCACTAGCTGGGACAGAGCTTCTCAGTATACAAATTAACAA 3120  
 Y A E T V H O D T S E V Y T D A P T S M K R F I R O Q L A M A E C S O T N H L E N -  
 ACCTCTCGCATATCAAAATCGCCCTCTATGTTTATTTATTTATTTACAGATATCTTACCTATCTACTTATAGCTTTCTGCTCAATATATCTCTGTCAAAATTTAAATATA 3240  
 T P M H I R N A P L N F F I Y T T D M I L P H L L I S F C V N I L K A I L N I -  
 ACTACAACTTTATACAGCTTCTATCGGCAAAATTTATATATGTTTTCGCAATCTTTTACTGTTTCGACAGCAAGCTTTAAAGCTATGTCAGAAATCAAGCTGATTATGTA 3360  
 S L V T X A S V N E X I L Y V L L C N I F S F G C R N F R A M S A N E U T Y V -

J2891 end cpo35  
 TTCTTATCTCGCTTTTATATCGTTTCAGATAAATATGTCCTCATTAAGCTATTAGCACTTATCAGATGTCATCATTTAGCGTGGCAACTAAATTAACAGACTGACAT 3480  
 FLIPVFIIIVLSIINCPIRLLLGLNRCSDDLGMGCTRNLTET -  
 AATACTACTCGGTATATAGACTATTTACTCAGACTATTAATGATTTTTGAAAAGAAAGCTGTTTTTAATCTTAAGCAAGAACTGGAATATCAGATATGACAGCACTGGAACTA 3490  
 TTTTACCTATCTTTTTCATATATTACCAATTATCTTGAAGTTTCTTTTAAAGCTGCTATCTGACTGCTCATCTTCAGAGCTAAGCGTCATATTTTATGTCGAAAAGCT 3500  
 - - - - -  
 GTTAAAGACTTAATCACTTATATATTTTATACAAATAGCTGAAGCAATGTTATCTATCAAGAAAGCTTACTGCTTCTGCTGGTGGCAAGCTGGCTGCTGCTGCTGCT 3640  
 - - - - -  
 KKKKKK  
 NKKVKKAVIFAAGACTGTRFLPAT -  
 - - - - -  
 AAGCTTTGCAAGAAAGAAATGCTGCATATGACACCCGCAAGCAATCATTTCTGCTCAAGCAAGCTTACTGCTGCGGTATCAAGATCTTACTGCTGAAGCACTAAGCT 3560  
 KALAKENLPIVDRPTINFVITEEALRSCTIEDILVVTGKSER -  
 TCTATCAAGATTATTTGATTCACTTTTCATTGCAATATAGCTTTAGAAAACAGCAAGACTGCAACTTTTAACTCACTTAACCAATCCACTCATATAACTACATTTCTGCT 4080  
 SIEDYFDSYTFELEYSLRKGCKNELLESVNESTOEVHFVR -  
 - - - - -  
 CAAAGTTCAGCACTGCTCTGCT 4200  
 QSPFRGLGDAYLQAKRSFVGGDDPFPVUNLCDDLDMDITOSTAV -  
 - - - - -  
 CTTTAAACAGCAATTTATGCTGATTACAGCAAGCAAGCAAGCTCAACTATCCGAGTAATCGCTTGATAGATCAAGACTTTCTCTTCTGCTGCTGCTGCTGCTGCTGCT 4320  
 PLTRLDHDDYHATQASTIAYHVPVATGCDVSSYGV:SPRLES -  
 - - - - -  
 AGTAATCCGCTCTATGCTGTTGATGCTTTTGTAGACAACCAAAACCAAGCAAGCGGCTACCCATTAGCTATTATTCGAGCTATCTACTTACTGCTGCTGCTGCTGCT 4440  
 SNGLYSVDAFVKPRPFEAPASFKLALIGRITLGPETGCGLE:SPRLES -  
 - - - - -  
 ACCCAAAAGCCAGCAGCAATGAATTCATTCAGCACTATTGATACATTAATAAGCAAGCAAGCTGTTTTCCGCTCAATTTCTGCGCAAACTTACAGCTCTGCTGCTAAG 4560  
 TORKPAGCHEIQDLTDAIDGTLNKTQSVFAREFVVGKATDVGDGR -  
 - - - - -  
 TTTTATTTTAAACACATCAATGCTATATGCTTTCACAGCTCTGCAAGCTAAAGCAGCTTTAAAAAATACCTTATTCAGCTTGGTAAACAACTGCAAGCTGACTGCTGCTCA 4680  
 FNFHRTSLOTALQHPQIKESLANYVIALKQLESLDCCS -  
 - - - - -  
 end cpo35  
 ACTGCAAGCTCAATTTCTATAGAAAGTATCAAAAATGCTTAATGTCGCTGATCTC 4740  
 S.E.N.L.

FIG. 6H

FIG. 6I

**FIG. 6J**

TAAATCCCTATTTCCAAATGCACAAATCGCTCTGTATACATTTCATATCATCCAGCACCTTAAGCGGTGCAGATCATCAGACAGAAATTTGCATCATGAAGAATTTGCAGCAAGTAATC  
 L N A Y F T C A G H K S S L I D F I S S S T L S R E N H T T R I V I N K H F D E Q S .  
 C T T C T C T A T C G C C C A T C A A A A A C T C G C C C A A T T T C C A A A G A G T A G C C C G A A A A T C T C G A T A A G C C C T T T A C T C A G A A C T T T A G C A C A C A C T C A C A C T C G C A A A T G C T C  
 L S Y R A N K M N H W R I F O K D S R K L S D K A F Y S R T F R O T L T F P R E I V  
 A A T A A A A C C T A G C T T C A G A C A C C C A C A T T A T A G C A G T T A A C C A C T C C T C T T T T C A C C T G C A A A A A A G A T C C C G A C C A T T C T T T G A C T T G T T A A C T A G C C C T T T  
 N K T L A F S D E L H Y I D D L Y O L L L F R L O E K N A D R F F D L L T Z E A F  
 C C C A C T A C T A A C A A G C C T T A G A A C T C C T T A C C A C C C G T C G C C C T A C A G A A A A T A C G T C A C A A A T C C C A T C C A G A T C C C C T A T T A C T A C C A A C C T G C A C C C A A T C A A A A C T  
 P O L N Q A F R T A L S T L V R Y R K Y V T N A I O N P Y S N S K L Z A I N K L  
 A T C A C C T T C T C A A A C C C A A C A T G T A G C T T A T C T T T T C T C A A A T C T  
 I S F L Q Y T Q T C R L C F S S S S

9/17

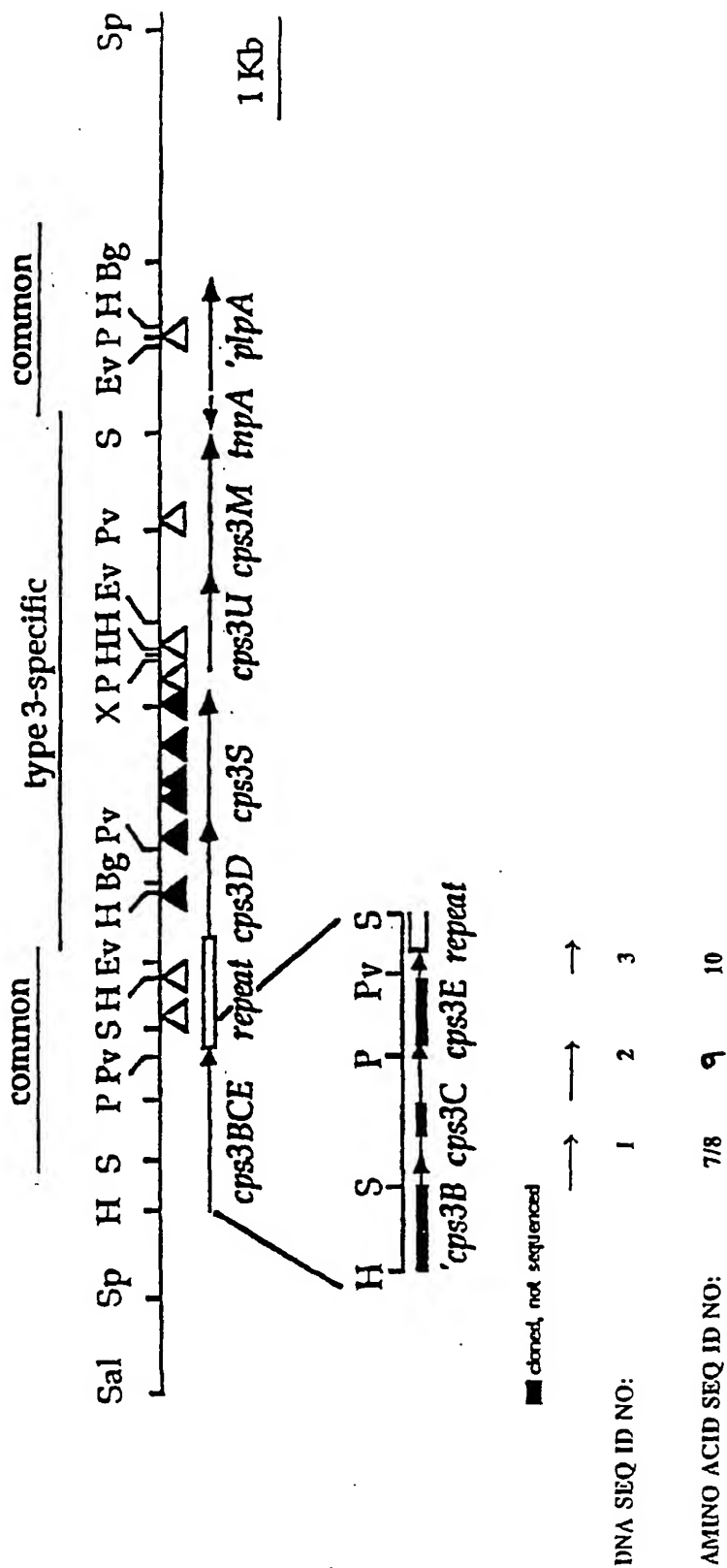


FIG. 7

10/17

1kbp

SEQ ID NO:6

SEQ ID NO:5

SEQ ID NO:4

TYPE 3 SPECIFIC

1 - 4951

1 - 1307

1 - 934

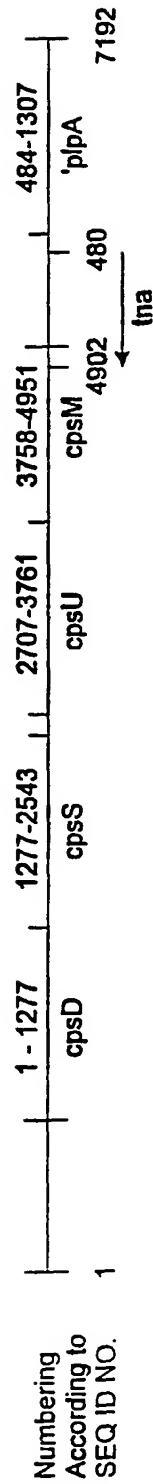


FIG 8

11/17

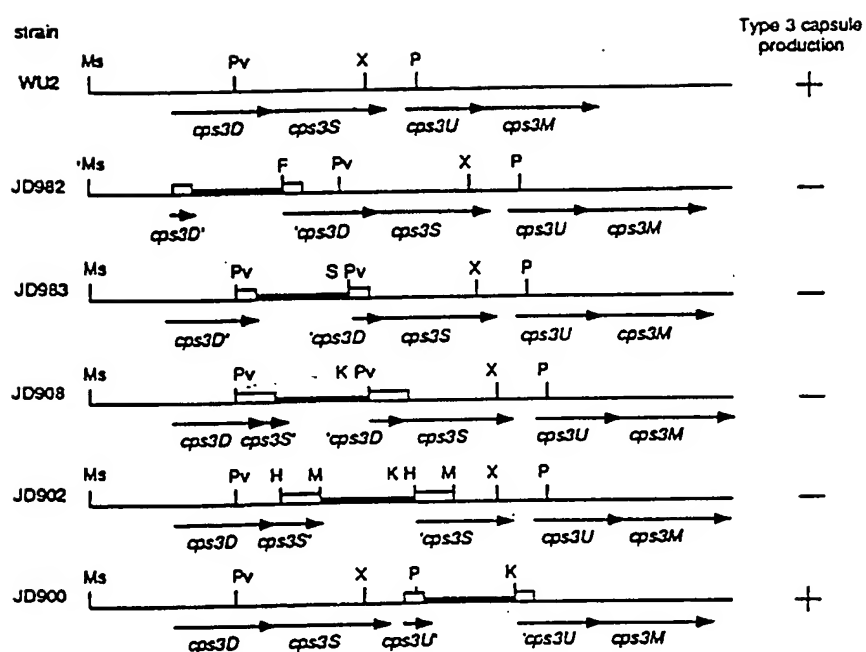


FIG. 9

12/17

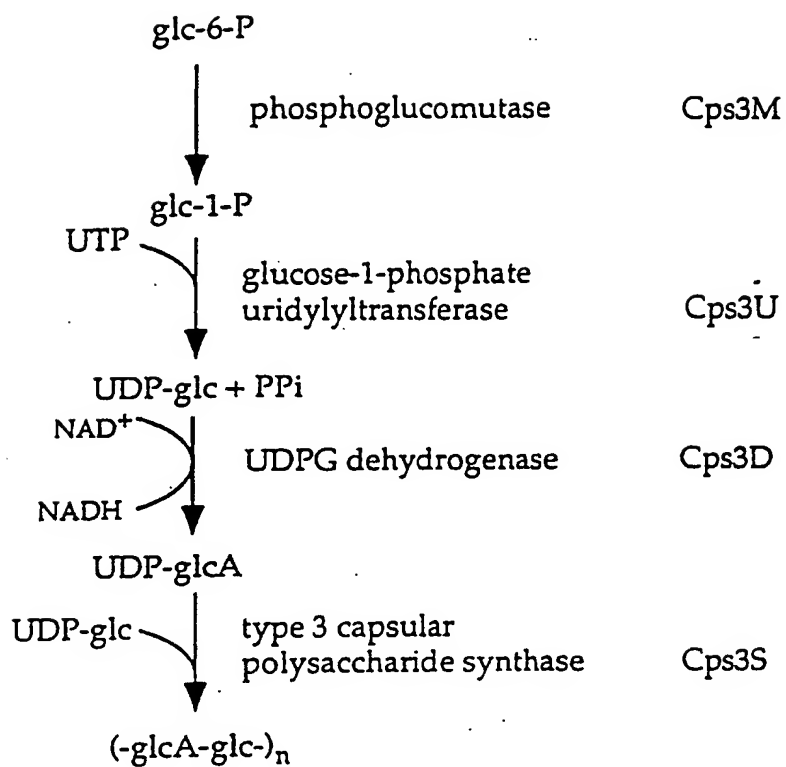


FIG. 10

13/17

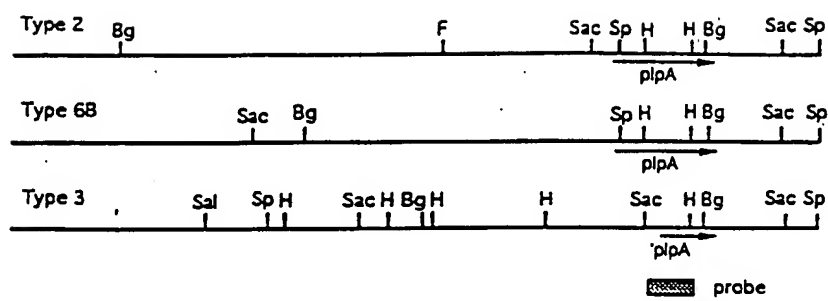


FIG. 11



14/17

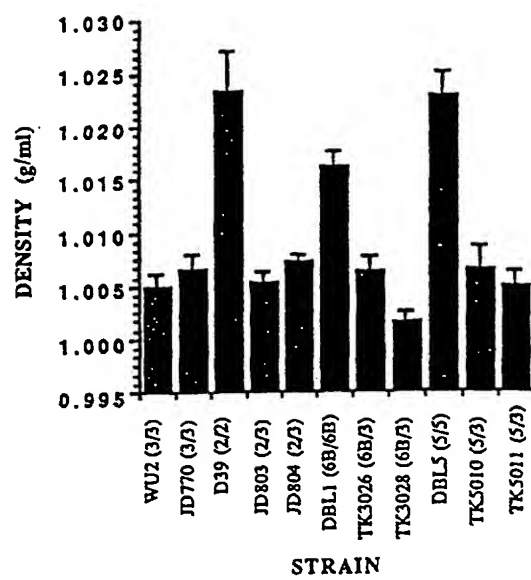


FIG. 12

15/17

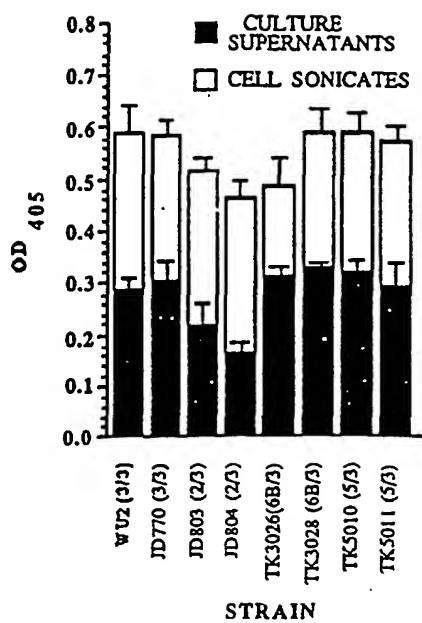


FIG. 13

16/17

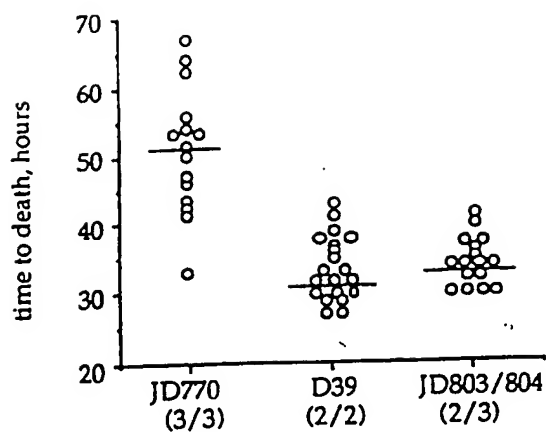


FIG. 14A

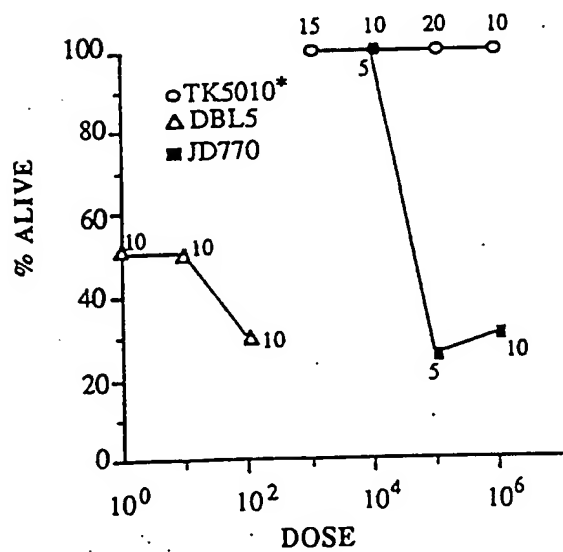


FIG. 14B

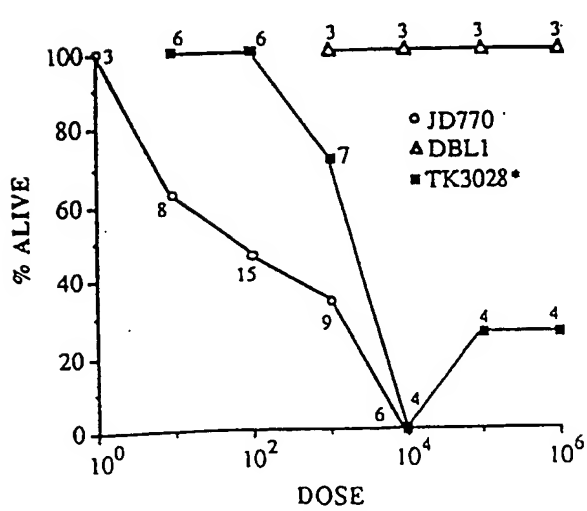


FIG. 14C

Cassette-type recombination

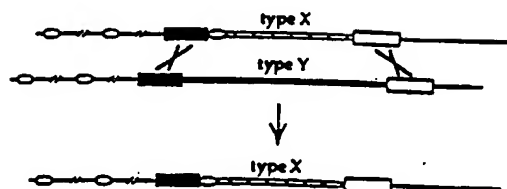


FIG. 15A

Binary encapsulation via homologous and illegitimate recombinations

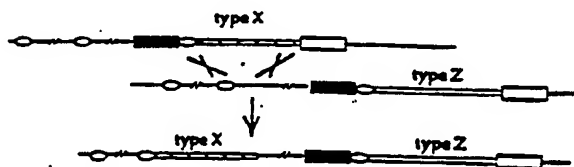


FIG. 15B

Binary encapsulation via a transposition-like event

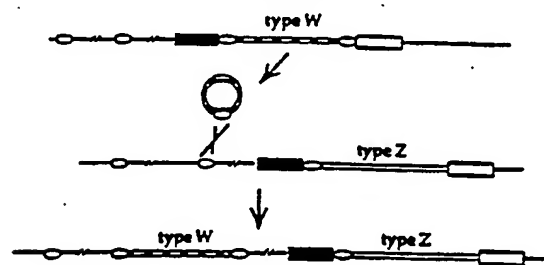


FIG. 15C

# INTERNATIONAL SEARCH REPORT

Inter. Application No  
PCT/US 95/06119

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 6   C12N15/31   C07K14/315   C12N15/52   C12N15/53   C12N15/60 C12N15/54   C12N15/61   C12Q1/68   C12P19/04   A61K39/09 C12N9/04   C12N9/88   C12N9/12   C12N9/90   C12N1/21					
According to International Patent Classification (IPC) or to both national classification and IPC					
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 6   C07K   C12N   C12P   C12Q					
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched					
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)					
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>					
Category *	Citation of document, with indication, where appropriate, of the relevant passages				Relevant to claim No.
X	MOL GEN GENET 239 (1-2). 1993. 188-195, GARCIA, E. ET AL. 'CLONING AND SEQUENCING OF A GENE INVOLVED IN THE SYNTHESIS OF THE CAPSULAR POLYSACCHARIDE OF STREPTOCOCCUS -PNEUMONIAE TYPE 3.' cited in the application see figures 2,4				1-11
A	MOLECULAR MICROBIOLOGY, vol. 9, no. 5, 1993 pages 1037-1050, PEARCE, B. ET AL. 'Genetic identification of exported proteins in Streptococcus pneumoniae' cited in the application see figure 4, Exp1				12-14
-/--					
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <span style="margin-left: 100px;"><input checked="" type="checkbox"/> Patent family members are listed in annex.</span>					
* Special categories of cited documents :					
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed			"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search  <div style="text-align: center; font-weight: bold;">11 September 1995</div>			Date of mailing of the international search report  <div style="text-align: center; font-weight: bold;">05.10.95</div>		
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016			Authorized officer  <div style="text-align: center; font-weight: bold;">Andres, S</div>		

## INTERNATIONAL SEARCH REPORT

Inter national Application No

PCT/US 95/06119

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	JOURNAL OF EXPERIMENTAL MEDICINE 181 (3). 973-983, March 1995 DILLARD, J. ET AL. 'Characterization of the Cassette Containing Genes for Type 3 Capsular Polysaccharide Biosynthesis in Streptococcus pneumoniae.' see the whole document ----	1-52
P,X	JOURNAL OF BACTERIOLOGY 176 (20). 6375-6383, October 1994 ARRECUBIETA, C. ET AL. 'Molecular characterization of cap3A, a gene from the operon required for the synthesis of the capsule of Streptococcus pneumoniae type 3: Sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome.' cited in the application see figures 1,2 ----	1-11
P,X	MOLECULAR MICROBIOLOGY 12 (6). 1994. 959-972, DILLARD, J. ET AL. 'Genetic and molecular characterization of capsular polysaccharide biosynthesis in Streptococcus pneumoniae type 3.' cited in the application see figure 4 ----	1-52
P,X	INFECTION AND IMMUNITY 62 (5). 1813-1819, December 1994 KELLY, T. ET AL. 'Effect of genetic switching of capsular type on virulence of Streptococcus pneumoniae.' cited in the application see figures 1,2; table 3 ----	1,2
P,X	WO-A-95 06732 (UNIV ROCKEFELLER ;MASURE H ROBERT (US); PEARCE BARBARA J (US); TUO) 9 March 1995 see example 3 see SEQ IDs 1 and 46 -----	12-14, 31,39

**information on patent family members**

**PCT/US 95/06119**

**Patent document  
cited in search report**

Publication date

Patent family member(s)

Publication date

**WO-A-9506732**

09-03-95

**AU-B- 7680994**

**22-03-95**